

A Parallel Douglas Rachford Algorithm for Minimizing ROF-like Functionals on Images with Values in Symmetric Hadamard Manifolds

Extended Version

Ronny Bergmann* Johannes Persch* Gabriele Steidl*

March 24, 2016

Abstract

We are interested in restoring images having values in a symmetric Hadamard manifold by minimizing a functional with a quadratic data term and a total variation like regularizing term. To solve the convex minimization problem, we extend the Douglas-Rachford algorithm and its parallel version to symmetric Hadamard manifolds. The core of the Douglas-Rachford algorithm are reflections of the functions involved in the functional to be minimized. In the Euclidean setting the reflections of convex lower semicontinuous functions are nonexpansive. As a consequence, convergence results for Krasnoselski-Mann iterations imply the convergence of the Douglas-Rachford algorithm. Unfortunately, this general results does not carry over to Hadamard manifolds, where proper convex lower semicontinuous functions can have expansive reflections. However, splitting our restoration functional in an appropriate way, we have only to deal with special functions namely, several distance-like functions and an indicator functions of a special convex sets. We prove that the reflections of certain distance-like functions on Hadamard manifolds are nonexpansive which is an interesting result on its own. Furthermore, the reflection of the involved indicator function is nonexpansive on Hadamard manifolds with constant curvature so that the Douglas-Rachford algorithm converges here.

Several numerical examples demonstrate the advantageous performance of the suggested algorithm compared to other existing methods as the cyclic proximal point algorithm or half-quadratic minimization. Numerical convergence is also observed in our experiments on the Hadamard manifold of symmetric positive definite matrices with the affine invariant metric which does not have a constant curvature.

1. Introduction

In the original paper [28], the Douglas-Rachford (DR) algorithm was proposed for the numerical solution of a partial differential equation, i.e., for solving systems of linear equations.

*Department of Mathematics, Technische Universität Kaiserslautern, Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany, {bergmann, persch, steidl}@mathematik.uni-kl.de.

It was generalized for finding a zero of the sum of two maximal monotone operators by Lions and Mercier [55], see also Passty's paper [61]. Eckstein and Bertsekas [29] examined the algorithm with under/over-relaxation and inexact inner evaluations. Gabay [35] considered problems of a special structure and showed that the DR algorithm applied to the dual problem results in the alternating direction method of multipliers (ADMM) introduced in [36, 39]. For relations between the DR algorithm and the ADMM we also refer to [40, 67]. In [82] it is shown that the ADMM is in some sense self-dual, i.e., it is not only equivalent to the DR algorithm applied to the dual problem, but also to the primal one. Recently, these algorithms were successfully applied in image processing mainly for two reasons: the functionals to minimize allow for simple proximal mappings within the method, and it turned out that the algorithms are highly parallelizable, see, e.g., [24]. Therefore the algorithms became one of the most popular ones in variational methods for image processing. Continued interest in DR iterations is also due to its excellent, but still myserious performance on various non-convex problems, see, e.g., [11, 12, 17, 30, 41, 45] and for recent progress on the convergence of ADMM methods for special non-convex problems [46, 54, 56, 75, 76, 81].

In this paper, we derive a parallel DR algorithm to minimize functionals on finite dimensional, symmetric Hadamard manifolds. The main ingredients of the algorithm are reflections of the functions involved in the functional we want to minimize. If these reflections are nonexpansive, then it follows from general convergence results of Krasnoselski-Mann iterations in CAT(0) spaces [48] that the sequence produced by the algorithm converges. Unfortunately, the well-known result in the Euclidean setting that reflections of convex lower semicontinuous functions are nonexpansive does not carry over to the Hadamard manifold setting, see [19, 32]. Up to now it was only proved that indicator functions of closed convex sets in Hadamard manifolds with constant curvature possess nonexpansive reflections [32]. In this paper, we prove that certain distance-like functions have nonexpansive reflections on general symmetric Hadamard manifolds. Such distance-like functions appear for example in the manifold-valued counterpart of the Rudin-Osher-Fatemi (ROF) model [66].

The ROF model is the most popular variational model for image restoration, in particular for image denoising. For real-valued images, its discrete, anisotropic penalized form is given by

$$\begin{aligned} D(u; f) + \alpha \text{TV}(u) &= \frac{1}{2} \|f - u\|_2^2 + \alpha \|\nabla u\|_1 \\ &= \frac{1}{2} \sum_{i,j} (f_{i,j} - u_{i,j})^2 + \alpha \sum_{i,j} (|u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|), \end{aligned} \quad (1)$$

where $f = (f_{i,j}) \in \mathbb{R}^{N,M}$ is an initial corrupted image and ∇ denotes the discrete gradient operator usually consisting of first order forward differences in vertical and horizontal directions. The first term is the *data fidelity term* $\mathcal{D}(u; f)$ measuring similarity between u and the given data f . The second term $\text{TV}(u)$ is the total variation (TV) type *regularizer* posing a small value of the first order differences in u . The regularization parameter $\alpha > 0$ steers the relation between both terms. The popularity of the model arises from the fact that its minimizer is a smoothed image that preserves important features such as edges. There is a close relation of the ROF model to PDE and wavelet approaches, see [70].

In various applications in image processing and computer vision the functions of interest take values in a Riemannian manifold. One example is diffusion tensor imaging where the data is given on the Hadamard manifold of positive definite matrices; see, e.g., [9, 21,

20, 22, 62, 69, 78, 80]. In the following we are interested in generalizations of ROF-like functionals to manifold-valued settings, more precisely to data having values in symmetric Hadamard manifolds. In [37, 38], the notion of the total variation of functions having their values on a manifold was investigated based on the theory of Cartesian currents. The first work which applies a TV approach of circle-valued data for image processing tasks is [71]. An algorithm for TV regularized minimization problems on Riemannian manifolds was proposed in [52]. There, the problem is reformulated as a multilabel optimization problem which is approached using convex relaxation techniques. Another approach to TV minimization for manifold-valued data which employs cyclic and parallel proximal point algorithms and does not require labeling and relaxation techniques, was given in [79]. A method which circumvents the direct work with manifold-valued data by embedding the matrix manifold in the appropriate Euclidean space and applying a back projection to the manifold was suggested in [65]. TV-like functionals on manifolds with higher order differences were handled in [8, 15, 16]. Finally we mention the relation to wavelet-type multiscale transforms which were handled, e.g., in [42, 43, 63, 74].

We will apply the parallel DR algorithm to minimize the ROF-like functional for images having values in a symmetric Hadamard manifolds. To this end we will split the functional in an appropriate way and show the convergence of the algorithm by examining the reflections of the involved distance-like functions. In the numerical part we show the very good performance of the proposed parallel DR algorithm for various symmetric Hadamard manifolds with and without constant curvature. In particular, we compare the algorithm with other algorithms existing in the literature, namely the cyclic proximal point algorithm [79] and a half-quadratic minimization method applied to a smoothed version of the ROF-like functional [14].

The outline of the paper is as follows: We start by recalling the DR algorithm and its parallel version in the Euclidean setting in Section 2. The generalization to symmetric Hadamard manifolds will follow the same path. In Section 3 we provide the notation and preliminaries in Hadamard manifolds which are required to understand our subsequent findings. The parallel DR algorithm on symmetric Hadamard manifolds is outlined in Section 4. We prove that the sequence produced by the algorithm converges to a minimizer of the functional. In Section 5 we show how the parallel DR algorithm can be applied to minimize a ROF-like functional which can be used for restoring images with values in symmetric Hadamard manifolds. Convergence of the algorithm is ensured if the reflections of the functions appearing in the functional are nonexpansive. For the ROF-like functional we have, due to an appropriate splitting, only to consider distance-like functions and an indicator function of a convex set. Section 6 contains the interesting result that reflections of certain distance-like functions on symmetric Hadamard manifolds are nonexpansive. In Section 7 we will see that indicator functions of closed convex sets have nonexpansive reflections on manifolds with constant curvature. Numerical examples are demonstrated in Section 8. These include manifolds such as the hyperbolic model space and the space of symmetric positive definite matrices. Comparisons with other algorithms to minimize an ROF-like functional on Hadamard manifolds are given. Conclusions are drawn in Section 9. The appendix provides material on symmetric positive definite matrices and hyperbolic spaces which are necessary for the implementation of the generalized parallel DR algorithm.

Algorithm 1 DR Algorithm for Real-Valued Data

Input: $t^{(0)} \in \mathbb{R}^n$, $\lambda_r \in [0, 1]$ with $\sum_{r \in \mathbb{N}} \lambda_r(1 - \lambda_r) = +\infty$, $\eta > 0$
 $r = 0$;
repeat
 $t^{(r+1)} = ((1 - \lambda_r) \text{Id} + \lambda_r \mathcal{R}_{\eta\varphi} \mathcal{R}_{\eta\psi})(t^{(r)})$;
 $r \rightarrow r + 1$;
until a stopping criterion is reached

2. Parallel DR Algorithm on Euclidean Spaces

We start by recalling the DR algorithm and its parallel form on Euclidean spaces. We will follow the same path for data in a Hadamard manifold in Section 4. The main ingredients of the DR algorithm are proximal mappings and reflections.

For $\eta > 0$ and a proper convex lower semicontinuous (lsc) function $\varphi: \mathbb{R}^n \rightarrow (-\infty, +\infty]$, the *proximal mapping* $\text{prox}_{\eta\varphi}$ reads

$$\text{prox}_{\eta\varphi}(x) := \arg \min_{y \in \mathbb{R}^n} \left\{ \frac{1}{2} \|x - y\|_2^2 + \eta\varphi(y) \right\},$$

see [59]. It is well-defined and unique. The *reflection* $R_p: \mathbb{R}^n \rightarrow \mathbb{R}^n$ at a point $p \in \mathbb{R}^n$ is given by

$$R_p(x) = 2p - x.$$

Further, let $\mathcal{R}_\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ denote the reflection operator at prox_φ , i.e.,

$$\mathcal{R}_\varphi(x) = 2\text{prox}_\varphi(x) - x.$$

We call this operator *reflection of the function* φ .

Given two proper convex lsc functions $\varphi, \psi: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ the DR algorithm aims to solve

$$\arg \min_{x \in \mathbb{R}^n} \{ \varphi(x) + \psi(x) \} \tag{2}$$

by the steps summarized in Algorithm 1. It is known that the DR algorithm converges for any proper convex lsc functions φ, ψ under mild assumptions. More precisely, we have the following theorem, see [55] or [10, Theorem 27.4].

Theorem 2.1. *Let $\varphi, \psi: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ be proper convex lsc functions such that $\text{ri}(\text{dom } \varphi) \cap \text{ri}(\text{dom } \psi) \neq \emptyset$. Assume that a solution of (2) exists. Let $\{\lambda_r\}_{r \in \mathbb{N}}$ fulfill $\sum_{r \in \mathbb{N}} \lambda_r(1 - \lambda_r) = +\infty$ and $\eta > 0$. Then the sequence $\{t^{(r)}\}_{r \in \mathbb{N}}$ generated by Algorithm 1 converges for any starting point $t^{(0)}$ to a point \hat{t} , and*

$$\hat{x} := \text{prox}_{\eta\psi}(\hat{t}) \tag{3}$$

is a solution of (2).

The DR algorithm can be considered as a special case of the *Krasnoselski-Mann iteration*

$$t^{(r+1)} = ((1 - \lambda_r) \text{Id} + \lambda_r T)(t^{(r)}) \tag{4}$$

with $T := \mathcal{R}_{\eta\varphi}\mathcal{R}_{\eta\psi}$. Since the concatenation and convex combination of nonexpansive operators is again nonexpansive, the proof is just based on the fact that the reflection operator $\mathcal{R}_{\eta\varphi}$ of a proper convex lsc function φ is nonexpansive. The latter will not remain true in the Hadamard manifold setting.

Remark 2.2. *The definition of the DR algorithm is dependent on the order of the operators $\mathcal{R}_{\eta\varphi}$ and $\mathcal{R}_{\eta\psi}$ although problem (2) itself is not. The order of the operators in the DR iterations was examined in [13]. The authors showed that $\mathcal{R}_{\eta\varphi}$ is an isometric bijection from the fixed point set of $\frac{1}{2}(\text{Id} + \mathcal{R}_{\eta\psi}\mathcal{R}_{\eta\varphi})$ to that of $\frac{1}{2}(\text{Id} + \mathcal{R}_{\eta\varphi}\mathcal{R}_{\eta\psi})$ with inverse $\mathcal{R}_{\eta\psi}$. For the effect of different orders of the operators in the ADMM algorithm see [83].*

In this paper, we are interested in multiple summands. We consider

$$\arg \min_{x \in \mathbb{R}^n} \left\{ \sum_{k=1}^K \varphi_k(x) \right\}, \quad (5)$$

where $\varphi_k: \mathbb{R}^n \rightarrow (-\infty, +\infty]$, $k = 1, \dots, K$, are proper convex lsc functions with $\bigcap_{k=1}^K \text{ri}(\text{dom } \varphi_k) \neq \emptyset$. The problem can be rewritten in the form (2) with only two summands, namely

$$\arg \min_{\mathbf{x} \in \mathbb{R}^{nK}} \{ \Phi(\mathbf{x}) + \iota_{\mathbf{D}}(\mathbf{x}) \}, \quad (6)$$

where $\Phi(\mathbf{x}) := \sum_{k=1}^K \varphi_k(x_k)$, $\mathbf{x} := (x_k)_{k=1}^K$, $x_k \in \mathbb{R}^n$, $k = 1, \dots, K$, and

$$\iota_{\mathbf{D}}(\mathbf{x}) := \begin{cases} 0 & \text{if } \mathbf{x} \in \mathbf{D}, \\ \infty & \text{otherwise,} \end{cases}$$

is the indicator function of

$$\mathbf{D} := \{ \mathbf{x} \in \mathbb{R}^{nK} : x_1 = \dots = x_K \in \mathbb{R}^n \}.$$

Since \mathbf{D} is a nonempty closed convex set, its indicator function is proper convex and lsc. Further, we have

$$\text{prox}_{\iota_{\mathbf{D}}}(\mathbf{x}) = \Pi_{\mathbf{D}}(\mathbf{x}) = \left(\frac{1}{K} \sum_{k=1}^K x_k, \dots, \frac{1}{K} \sum_{k=1}^K x_k \right) \in \mathbb{R}^{nK} \quad (7)$$

By $\Pi_{\mathbf{D}}$ we denote the orthogonal projection onto \mathbf{D} . If (6) has a solution, then we can apply the DR Algorithm 1 to the special setting in (6) and obtain Algorithm 2. We call this algorithm *parallel DR algorithm*. In the literature, it is also known as product version of the DR algorithm [18], parallel proximal point algorithm [24] or, in a slightly different version, as parallel splitting algorithm [10, Proposition 27.8]. For a stochastic version of the algorithm we refer to [25].

By Theorem 2.1, the convergence of the iterates of Algorithm 2 to a point $\hat{\mathbf{t}} = (\hat{t}_1, \dots, \hat{t}_K)^T \in \mathbb{R}^{nK}$ is ensured under the assumptions of Theorem 2.1. Further, by (3) and (7), we obtain that

$$\hat{x} := \frac{1}{K} \sum_{k=1}^K \hat{t}_k$$

is a solution of (5) resp. (6).

Finally, we want to mention the so-called cyclic DR algorithm.

Algorithm 2 Parallel DR Algorithm for Real-Valued Data

Input: $\mathbf{t}^{(0)} \in \mathbb{R}^{nK}$, $\lambda_r \in [0, 1]$ with $\sum_{r \in \mathbb{N}} \lambda_r(1 - \lambda_r) = +\infty$, $\eta > 0$
 $r = 0$;
repeat
 $\mathbf{t}^{(r+1)} = ((1 - \lambda_r) \text{Id} + \lambda_r \mathcal{R}_{\eta\Phi} \mathcal{R}_{\iota_D})(\mathbf{t}^{(r)})$;
 $r \rightarrow r + 1$;
until a stopping criterion is reached

Remark 2.3. (*Cyclic DR Algorithm*) We consider the cyclic DR algorithm for (5). Let $S_k := \mathcal{R}_{\eta\varphi_{k+1}} \mathcal{R}_{\eta\varphi_k}$, $k = 1, \dots, K-1$, and $S_K := \mathcal{R}_{\eta\varphi_1} \mathcal{R}_{\eta\varphi_K}$. Starting with $t^{(0)} \in \mathbb{R}^n$ we compute

$$t^{(r+1)} = T_{[\varphi_1, \dots, \varphi_K]} t^{(r)},$$

where

$$T_{[\varphi_1, \dots, \varphi_K]} := \frac{1}{2}(\text{Id} + S_K) \circ \frac{1}{2}(\text{Id} + S_{K-1}) \circ \dots \circ \frac{1}{2}(\text{Id} + S_2) \circ \frac{1}{2}(\text{Id} + S_1).$$

Applying the operator $T_{[\varphi_1, \dots, \varphi_K]}$ can be written again in the form of a Krasnoselski-Mann iteration (4) with operator $\mu \text{Id} + (1 - \mu)T$ where $\mu = (\frac{1}{2})^K$ and T is the convex combination

$$\frac{1}{2^K - 1} (S_1 + \dots + S_K + S_2 \circ S_1 + \dots + S_K \circ S_{K-1} + \dots + S_K \circ \dots \circ S_1).$$

Thus it can be shown that the sequence $\{t^{(r)}\}_{r \in \mathbb{N}}$ converges to some \hat{t} . However, only for indicator functions $\varphi_k := \iota_{C_k}$ of closed, convex sets $C_k \neq \emptyset$ it was proved in [18] that $\hat{x} = \text{prox}_{\iota_{C_k}}(\hat{t}) = \Pi_{C_k}(\hat{t})$ is a solution of (5). In other words, the algorithm finds an element of $\cap_{k=1}^K C_k$. For non-indicator functions φ_k , there are to the best of our knowledge no similar results, i.e., the algorithm converges but the meaning of the limit is not clear. The same holds true for an averaged version of the DR algorithm, see [18].

3. Preliminaries on Hadamard Manifolds

A complete metric space (\mathcal{H}, d) is called a *Hadamard space* if every two points x, y are connected by a geodesic and the following condition holds true

$$d(x, v)^2 + d(y, w)^2 \leq d(x, w)^2 + d(y, v)^2 + 2d(x, y)d(v, w), \quad (8)$$

for any $x, y, v, w \in \mathcal{H}$. Inequality (8) implies that Hadamard spaces have nonpositive curvature [3, 64]. In this paper we restrict our attention to Hadamard spaces which are at the same time finite dimensional Riemannian manifolds with geodesic distance d .

Let $T_x \mathcal{H}$ be the tangent space at x on \mathcal{H} . By $\exp_x: T_x \mathcal{H} \rightarrow \mathcal{H}$ we denote the exponential map at x . Then we have $\exp_x \xi = \gamma_{x, \xi}(1) = y$, where $\gamma_{x, \xi}: [0, 1] \rightarrow \mathcal{H}$ is the unique geodesic starting at x in direction ξ , i.e., $\gamma_{x, \xi}(0) = x$ and $\dot{\gamma}_{x, \xi}(0) = \xi$. Likewise the geodesic starting at x and reaching y at time $t = 1$ is denoted by $\gamma_{x, y}$. Its inverse $\log_x: \mathcal{H} \rightarrow T_x \mathcal{H}$ is given by $\log_x y = \xi$, where $\gamma_{x, \xi}(1) = y$.

Hadamard spaces have the nice feature that they resemble convexity properties from Euclidean spaces. A set $\mathcal{C} \subseteq \mathcal{H}$ is *convex*, if for any $x, y \in \mathcal{H}$ the geodesic $\gamma_{x, y}$ lies in \mathcal{C} . The intersection of an arbitrary family of convex closed sets is itself a convex closed set.

A function $\varphi: \mathcal{H} \rightarrow (-\infty, +\infty]$ is called *convex* if for any $x, y \in \mathcal{H}$ the function $\varphi \circ \gamma_{x,y}$ is convex, i.e., we have

$$\varphi(\gamma_{x,y}(t)) \leq t\varphi(\gamma_{x,y}(0)) + (1-t)\varphi(\gamma_{x,y}(1)), \text{ for all } t \in [0, 1].$$

The function φ is *strictly convex* if the strict inequality holds true for all $0 < t < 1$, and *strongly convex* with parameter $\kappa > 0$ if for any $x, y \in \mathcal{H}$ and all $t \in [0, 1]$ we have

$$\varphi(\gamma_{x,y}(t)) \leq t\varphi(\gamma_{x,y}(0)) + (1-t)\varphi(\gamma_{x,y}(1)) - \kappa t(1-t)d(\gamma_{x,y}(0), \gamma_{x,y}(1)).$$

The distance in an Hadamard space fulfills the following properties:

(D1) $d: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_{\geq 0}$ and $d^2: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_{\geq 0}$ are convex, and

(D2) $d^2(\cdot, y): \mathcal{H} \rightarrow \mathbb{R}_{\geq 0}$ is strongly convex with $\kappa = 1$.

Concerning minimizers of convex functions the next theorem summarizes some basic facts which can be found, e.g., in [7, Lemma 2.2.19] and [7, Proposition 2.2.17].

Theorem 3.1. *For proper convex lsc functions $\varphi: \mathcal{H} \rightarrow (-\infty, +\infty]$ we have:*

- i) *If $\varphi(x) \rightarrow +\infty$ whenever $d(x, x_0) \rightarrow +\infty$ for some $x_0 \in \mathcal{H}$, then φ has a minimizer.*
- ii) *If φ is strongly convex, then there exists a unique minimizer of φ .*

The *subdifferential* of $\varphi: \mathcal{H} \rightarrow (-\infty, +\infty]$ at $x \in \text{dom } \varphi$ is defined by

$$\partial\varphi(x) := \{\xi \in T_x\mathcal{H}: \varphi(y) \geq \varphi(x) + \langle \xi, \dot{\gamma}_{x,y}(0) \rangle \text{ for all } y \in \text{dom } \varphi\},$$

see, e.g., [53] or [73] for finite functions φ . For any $x \in \text{int}(\text{dom } \varphi)$, the subdifferential is a nonempty convex and compact set in $T_x\mathcal{H}$. If the Riemannian gradient $\nabla\varphi(x)$ of φ in $x \in \mathcal{H}$ exists, then $\partial\varphi(x) = \{\nabla\varphi(x)\}$. Further we see from the definition that $x \in \mathcal{H}$ is a global minimizer of φ if and only if $0 \in \partial\varphi(x)$. The following theorem was proved in [53] for general finite dimensional Riemannian manifolds.

Theorem 3.2. *Let $\varphi, \psi: \mathcal{H} \rightarrow (-\infty, +\infty]$ be proper and convex. Let $x \in \text{int}(\text{dom } \varphi) \cap \text{dom } \psi$. Then we have the subdifferential sum rule*

$$\partial(\varphi + \psi)(x) = \partial\varphi(x) + \partial\psi(x).$$

In particular, for a convex function φ and a nonempty convex set \mathcal{C} such that $\mathcal{C} \cap \text{dom } \varphi \neq \emptyset$ is convex, it follows for $x \in \text{int}(\text{dom } \varphi) \cap \mathcal{C}$ that

$$\partial(\varphi + \iota_{\mathcal{C}}) = \partial\varphi(x) + N_{\mathcal{C}}(x),$$

where the normal cone $N_{\mathcal{C}}(x)$ of \mathcal{C} at $x \in \mathcal{C}$ is defined by

$$N_{\mathcal{C}}(x) := \{\xi \in T_x\mathcal{H}: \langle \xi, \log_x c \rangle \leq 0 \text{ for all } c \in \mathcal{C}\}.$$

We will deal with the product space \mathcal{H}^n of a Hadamard manifold \mathcal{H} with distance

$$d(x, y) := \left(\sum_{k=1}^n d(x_k, y_k)^2 \right)^{\frac{1}{2}}$$

which is again a Hadamard manifold.

For $\eta > 0$ and a proper convex lsc function $\varphi: \mathcal{H} \rightarrow [-\infty, +\infty]$, the *proximal mapping*

$$\text{prox}_{\eta\varphi}(x) := \arg \min_{y \in \mathcal{H}^n} \left\{ \frac{1}{2} d(x, y)^2 + \eta\varphi(y) \right\}$$

exists and is uniquely determined, see [47, 57].

To introduce a DR algorithm we have to define reflections on manifolds. A mapping $R_p: \mathcal{M} \rightarrow \mathcal{M}$ on a Riemannian manifold \mathcal{M} is called *geodesic reflection* at $p \in \mathcal{M}$, if

$$R_p(p) = p, \quad \text{and} \quad D_p(R_p) = -I,$$

where $D_p(R_p)$ denotes the differential of R_p at $p \in \mathcal{M}$. For any x, p on a Hadamard manifold \mathcal{H}^n we can write the reflection as

$$R_p(x) = \exp_p(-\log_p x).$$

The *reflection of a proper convex lsc function* $\varphi: \mathcal{H}^n \rightarrow (-\infty, +\infty]$ is given by $\mathcal{R}_\varphi: \mathcal{H}^n \rightarrow \mathcal{H}^n$ with

$$\mathcal{R}_\varphi(x) = \exp_{\text{prox}_\varphi(x)}(-\log_{\text{prox}_\varphi(x)}(x)).$$

Finally a connected Riemannian manifold \mathcal{M} is called (globally) *symmetric*, if the geodesic reflection at any point $p \in \mathcal{M}$ is an isometry of \mathcal{M} , i.e., for all $x, y \in \mathcal{M}$ we have

$$d(R_p(x), R_p(y)) = d(x, y).$$

Examples are spheres, connected compact Lie groups, Grassmannians, hyperbolic spaces and symmetric positive definite matrices. The latter two manifolds are symmetric Hadamard manifolds. For more information we refer to [31, 44].

4. Parallel DR Algorithm on Hadamard Manifolds

In this section, we generalize the parallel DR algorithm to Hadamard manifolds. Given two proper convex lsc functions $\varphi, \psi: \mathcal{H}^n \rightarrow (-\infty, +\infty]$ the DR algorithm aims to solve

$$\arg \min_{x \in \mathcal{H}^n} \{ \varphi(x) + \psi(x) \}. \quad (9)$$

Adapting Algorithm 1 to the manifold-valued setting yields Algorithm 3. If the iterates of Algorithm 3 converge to a fixed point $\hat{t} \in \mathcal{H}^n$, then we will see in the next theorem that $\hat{x} := \text{prox}_{\eta\psi}(\hat{t})$ is a solution of (9).

Theorem 4.1. *Let \mathcal{H} be a Hadamard manifold and $\varphi, \psi: \mathcal{H}^n \rightarrow \mathbb{R}$ proper, convex, lsc functions such that $\text{int}(\text{dom } \varphi) \cap \text{dom } \psi \neq \emptyset$. Assume that a solution of (9) exists. Then, for each solution \hat{x} of (9) there exists a fixed point \hat{t} of $\mathcal{R}_{\eta\varphi}\mathcal{R}_{\eta\psi}$ such that*

$$\hat{x} = \text{prox}_{\eta\psi} \hat{t}.$$

Conversely, if a fixed point \hat{t} of $\mathcal{R}_{\eta\varphi}\mathcal{R}_{\eta\psi}$ exists, then \hat{x} defined as above is a solution of (9).

Algorithm 3 DR Algorithm for Data in Hadamard Manifolds

Input: $t^{(0)} \in \mathcal{H}^n$, $\lambda_r \in [0, 1]$ with $\sum_{r \in \mathbb{N}} \lambda_r(1 - \lambda_r) = +\infty$, $\eta > 0$
 $r = 0$;
repeat
 $s^{(r)} = \mathcal{R}_{\eta\phi} \mathcal{R}_{\eta\psi}(t^{(r)})$
 $t^{(r+1)} = \gamma_{\widehat{t^{(r)}, s^{(r)}}}(\lambda_r)$
 $r \rightarrow r + 1$;
until a stopping criterion is reached

The proof is given in Appendix A.

Next we are interested in the minimization of functionals containing multiple summands,

$$\arg \min_{x \in \mathcal{H}^n} \left\{ \sum_{k=1}^K \varphi_k(x) \right\}, \quad (10)$$

where $\varphi_k: \mathcal{H}^n \rightarrow (-\infty, +\infty]$, $k = 1, \dots, K$, are proper convex lsc functions. As in the Euclidean case the problem can be rewritten as

$$\arg \min_{x \in \mathcal{H}^{nK}} \{ \Phi(\mathbf{x}) + \iota_D(\mathbf{x}) \},$$

where $\Phi(\mathbf{x}) := \sum_{k=1}^K \varphi_k(x_k)$, $\mathbf{x} := (x_k)_{k=1}^K$, and

$$D := \{ \mathbf{x} \in \mathcal{H}^{nK} : x_1 = \dots = x_n \in \mathcal{H}^n \}.$$

Obviously, D is a nonempty, closed convex set so that its indicator function is proper convex and lsc, see [7, p. 37]. Further, we have for any $\eta > 0$ and $\mathbf{x} = (x_1, \dots, x_K)^T \in \mathcal{H}^{nK}$ that

$$\text{prox}_{\iota_D}(\mathbf{x}) = \Pi_D(\mathbf{x}) = \left(\arg \min_{x \in \mathcal{H}^n} \sum_{k=1}^K d(x_k, x)^2, \dots, \arg \min_{x \in \mathcal{H}^n} \sum_{k=1}^K d(x_k, x)^2 \right) \in \mathcal{H}^{nK}.$$

The minimizer of the sum is a so-called *Karcher mean*, Fréchet mean or Riemannian center of mass [49]. It can be efficiently computed on Riemannian manifolds using a gradient descent algorithm as investigated in [2] or on Hadamard manifolds by employing, e.g., a cyclic proximal point algorithm, see [6].

The parallel DR algorithm for data in a symmetric Hadamard manifold is given by Algorithm 4. If the sequence $\{\mathbf{t}^{(r)}\}_{r \in \mathbb{N}}$ converges to some $\hat{\mathbf{t}} = (\hat{t}_1, \dots, \hat{t}_K)^T$, then we know by Theorem 4.1 that

$$\hat{x} := \arg \min_{x \in \mathcal{H}^n} \sum_{k=1}^K d(\hat{t}_k, x)^2$$

is a solution of (10).

Remark 4.2. (*Cyclic DR Algorithm*) Since the cyclic DR iterations can be written similarly to Remark 2.3 as a Krasnoselski-Mann iteration we have for nonexpansive operators $R_{\eta\varphi_k}$, that cyclic DR algorithm converges to some fixed point. However, the relation of the fixed point to a solution of the minimization problem is, as in the Euclidean case with non indicator functions, completely unknown.

Algorithm 4 Parallel DR Algorithm for Data in Symmetric Hadamard Manifolds

Input: $\mathbf{t}^{(0)} \in \mathcal{H}^{nK}$, $\lambda_r \in [0, 1]$ with $\sum_{r \in \mathbb{N}} \lambda_r(1 - \lambda_r) = +\infty$, $\eta > 0$
 $r = 0$;
repeat
 $\mathbf{s}^{(r)} = \mathcal{R}_{\eta\Phi} \mathcal{R}_{\iota_D}(\mathbf{t}^{(r)})$
 $\mathbf{t}^{(r+1)} = \gamma_{\widehat{\mathbf{t}^{(r)}, \mathbf{s}^{(r)}}}(\lambda_r)$
 $r \rightarrow r + 1$;
until a stopping criterion is reached

It remains to examine under which conditions the sequence $\{\mathbf{t}^{(r)}\}_{r \in \mathbb{N}}$ converges. Setting $T := \mathcal{R}_{\eta\Phi} \mathcal{R}_{\eta\psi}$, the DR iteration can be seen as a *Krasnoselski–Mann iteration* given by

$$t^{(r+1)} := \gamma_{\widehat{t^{(r)}, T(t^{(r)})}}(\lambda_r). \quad (11)$$

The following theorem on the convergence of Krasnoselski–Mann iterations on Hadamard spaces was proved in [48], see also [7, Theorem 6.2.1].

Theorem 4.3. *Let (\mathcal{H}, d) be a finite dimensional Hadamard space and $T: \mathcal{H} \rightarrow \mathcal{H}$ a nonexpansive mapping with nonempty fixed point set. Assume that $(\lambda_r)_{r \in \mathbb{N}}$ satisfies $\sum_{r \in \mathbb{N}} \lambda_r(1 - \lambda_r) = \infty$. Then the sequence $\{t^{(r)}\}_{k \in \mathbb{N}}$ generated by the Krasnoselski–Mann iterations (11) converges for any starting point $t^{(0)} \in \mathcal{H}$ to some fixed point \hat{t} of T .*

Hence, if the operator $T = \mathcal{R}_{\eta\Phi} \mathcal{R}_{\iota_D}$ is nonexpansive the convergence of the parallel DR algorithm to some fixed point is ensured. Obviously, this is true if $\mathcal{R}_{\eta\Phi_k}$, $k = 1, \dots, K$ and \mathcal{R}_{ι_D} are nonexpansive. Unfortunately, the result from the Euclidean space that proper convex lsc functions produce nonexpansive reflections does not carry over the Hadamard manifold setting. However, we show in the following sections that the parallel DR algorithm is well suited for finding the minimizer of the ROF-like functional on images with values in a symmetric Hadamard manifold.

5. Application of the DR Algorithm to ROF-like Functionals

One of the most frequently used variational models for restoring noisy images or images with missing pixels is the ROF model [66]. In a discrete, anisotropic setting this model is given by (1). A generalization of the ROF model to images having cyclic values was proposed in [71] and to images with pixel values in a manifolds in [52, 79]. The model looks as follows: let $\mathcal{G} := \{1, \dots, N\} \times \{1, \dots, M\}$ be the image grid and $\emptyset \neq \mathcal{V} \subseteq \mathcal{G}$. Here \mathcal{V} denotes the set of available, in general noisy pixels. In particular, we have $\mathcal{V} = \mathcal{G}$ in the case of no missing pixels. Given an image $f: \mathcal{V} \rightarrow \mathcal{H}$ which is corrupted by noise or missing pixels, we want to restore the original image $u_0: \mathcal{G} \rightarrow \mathcal{H}$ as a minimizer of the functional

$$\begin{aligned} \mathcal{E}(u) &= \mathcal{D}(u; f) + \alpha \text{TV}(u) \\ &:= \frac{1}{2} \sum_{(i,j) \in \mathcal{V}} d(f_{i,j}, u_{i,j})^2 + \alpha \left(\sum_{(i,j) \in \mathcal{G}} d(u_{i,j}, u_{i+1,j}) + \sum_{(i,j) \in \mathcal{G}} d(u_{i,j}, u_{i,j+1}) \right) \end{aligned} \quad (12)$$

where $\alpha > 0$ and we assume mirror boundary conditions, i.e., $d(u_{i,j}, u_{i+1,j}) = 0$ if $(i+1, j) \notin \mathcal{G}$ and $d(u_{i,j}, u_{i,j+1}) = 0$ if $(i, j+1) \notin \mathcal{G}$. By property (D1) the functional \mathcal{E} is convex. If $\mathcal{V} \neq \emptyset$ then, regarding that the TV regularizer has only constant images in its kernel, we see by Theorem 3.1 i) that \mathcal{E} has a global minimizer. If $\mathcal{V} = \mathcal{G}$, then, by (D2) the functional is strongly convex and has a unique minimizer by Theorem 3.1 ii).

Reordering the image columnwise into a vector of length $n := NM$ we can consider f and u as elements in the product space \mathcal{H}^n which fits into the notation of the previous section. However, it is more convenient to keep the 2D formulation here. Denoting the data fidelity term by

$$\varphi_1(u) = D(u; f) = \sum_{(i,j) \in \mathcal{V}} \frac{1}{2} d(f_{i,j}, u_{i,j})^2$$

and splitting the regularizing term $\text{TV}(u)$ by grouping the odd and even indices with respect to both image dimensions as

$$\begin{aligned} \alpha \text{TV}(u) &= \alpha \sum_{\nu_1=0}^1 \sum_{i,j=1}^{\lfloor \frac{N-\nu_1}{2} \rfloor, M} d(u_{2i-1+\nu_1, j}, u_{2i+\nu_1, j}) \\ &\quad + \alpha \sum_{\nu_2=0}^1 \sum_{i,j=1}^N, \lfloor \frac{M-\nu_2}{2} \rfloor d(u_{i, 2j-1+\nu_2}, u_{i, 2j+\nu_2}) \\ &= \sum_{k=2}^5 \varphi_k(u) \end{aligned}$$

our functional becomes

$$\mathcal{E}(u) = \sum_{k=1}^5 \varphi_k(u). \quad (13)$$

This has exactly the form (10) with $K = 5$. We want to apply the parallel DR algorithm. To this end we have to compute the proximal mappings of the φ_k , $k = 1, \dots, K$. The reason for the above special splitting is that every pixel appears in each functional φ_k at most in one summand so that the proximal values can be computed componentwise. More precisely, we have for φ_1 and $x \in \mathcal{H}^{N,M}$ that

$$\begin{aligned} \text{prox}_{\eta\varphi_1}(x) &= \arg \min_{u \in \mathcal{H}^{N,M}} \left\{ \frac{1}{2} \sum_{(i,j) \in \mathcal{G}} d(x_{i,j}, u_{i,j})^2 + \frac{\eta}{2} \sum_{(i,j) \in \mathcal{V}} d(f_{i,j}, u_{i,j})^2 \right\} \\ &= \left(\text{prox}_{\eta g_{i,j}}(x_{i,j}) \right)_{(i,j) \in \mathcal{G}}, \end{aligned}$$

where $g_{i,j}: \mathcal{H} \rightarrow \mathbb{R}$ is defined by

$$g_{i,j} := \frac{1}{2} d(f_{i,j}, \cdot)^2 \quad (14)$$

if $(i, j) \in \mathcal{V}$ and $g_{i,j} := 0$ otherwise.

For φ_2 we get

$$\begin{aligned} \text{prox}_{\eta\varphi_2}(x) &= \arg \min_{u \in \mathcal{H}^{N,M}} \sum_{i,j=1}^{N,M} \frac{1}{2} d(x_{i,j}, u_{i,j})^2 + \eta \sum_{i,j=1}^{\lfloor \frac{N}{2} \rfloor, M} d(u_{2i-1,j}, u_{2i,j}) \\ &= \arg \min_{u \in \mathcal{H}^{N,M}} \sum_{i,j=1}^{\lfloor \frac{N}{2} \rfloor, M} \left(\frac{1}{2} d(x_{2i-1,j}, u_{2i-1,j})^2 + \frac{1}{2} d(x_{2i,j}, u_{2i,j})^2 + \eta d(u_{2i-1,j}, u_{2i,j}) \right) \\ &= \left(\text{prox}_{\eta G}(x_{2i-1,j}, x_{2i,j}) \right)_{i,j=1}^{\lfloor \frac{N}{2} \rfloor, M}, \end{aligned}$$

where $G: \mathcal{H}^2 \rightarrow \mathbb{R}$ is given by

$$G := d(\cdot, \cdot). \quad (15)$$

Similarly the proximal mappings of φ_k , $k = 3, 4, 5$, can be computed for pairwise components.

The proximal mappings of g and G , and consequently of our functions φ_k , $k = 1, \dots, K$, are given analytically, see [33, 79].

Lemma 5.1. *Let (\mathcal{H}, d) be an Hadamard manifold, $\eta > 0$ and $a \in \mathcal{H}$.*

i) *The proximal mappings of $g := \frac{1}{\nu} d(a, \cdot)^\nu$, $\nu \in \{1, 2\}$, are given by*

$$p_x = \text{prox}_{\eta g}(x) = \gamma_{x,a}(\hat{s})$$

with

$$\hat{s} := \begin{cases} \min \left\{ \frac{\eta}{d(x,a)}, 1 \right\} & \text{for } \nu = 1, \\ \frac{\eta}{1+\eta} & \text{for } \nu = 2. \end{cases}$$

ii) *The proximal mappings of $G := d(\cdot, \cdot)^\nu$, $\nu \in \{1, 2\}$, are given for $x = (x_0, x_2)$ by*

$$p_x = \text{prox}_{\eta G}(x) = (\gamma_{x_0, x_1}(\hat{s}), \gamma_{x_1, x_0}(\hat{s})),$$

with

$$\hat{s} := \begin{cases} \min \left\{ \frac{\eta}{d(x_0, x_1)}, \frac{1}{2} \right\} & \text{for } \nu = 1, \\ \frac{\eta}{1+2\eta} & \text{for } \nu = 2. \end{cases}$$

It follows that the reflection of φ_1 is just given by the componentwise reflections of the $g_{i,j}$, i.e., for $x \in \mathcal{H}^{N,M}$ we have

$$\mathcal{R}_{\varphi_1}(x) = (\mathcal{R}_{g_{i,j}}(x_{i,j}))_{i,j \in \mathcal{G}}.$$

Similarly, the reflection of φ_k , $k = 2, \dots, 5$, are determined by pairwise reflections of G , e.g.,

$$\mathcal{R}_{\varphi_2}(x) = (\mathcal{R}_G(x_{2i-1,j}, x_{2i,j}))_{i,j=1}^{\lfloor \frac{N}{2} \rfloor, M}.$$

Based on these considerations and Theorem 4.3 we see immediately that the parallel DR algorithm 4 applied to the special splitting (13) of the ROF-like functional on images with values in \mathcal{H} converges if the reflections $\mathcal{R}_{g_{i,j}}$, \mathcal{R}_G and \mathcal{R}_{ι_D} are nonexpansive. This is covered in the next two sections.

6. Reflections of Functions Related to Distances

In this section we will show that reflections with respect to the distance-like functions g and G in (14) and (15) are indeed nonexpansive which is an interesting result on its own.

We will need the following auxiliary lemma.

Lemma 6.1. *Let (\mathcal{H}, d) be an Hadamard space. Then for $x_0, x_1, y_0, y_1 \in \mathcal{H}$ it holds*

$$0 \leq d^2(x_0, y_0) + d^2(x_1, y_1) + d^2(x_0, x_1) + d^2(y_0, y_1) - d^2(x_0, y_1) - d^2(x_1, y_0). \quad (16)$$

For $s, t \in [0, 1]$ we have

$$\begin{aligned} d^2(\gamma_{x_0, x_1}(s), \gamma_{y_0, y_1}(t)) &\leq s t d^2(x_1, y_1) + (1-s)(1-t) d^2(x_0, y_0) \\ &\quad + (1-s)t d^2(x_0, y_1) + s(1-t) d^2(x_1, y_0) \\ &\quad - s(1-s) d^2(x_0, x_1) - t(1-t) d^2(y_0, y_1). \end{aligned} \quad (17)$$

For $s = t$, this becomes

$$\begin{aligned} d^2(\gamma_{x_0, x_1}(s), \gamma_{y_0, y_1}(s)) &\leq s^2 d^2(x_1, y_1) + (1-s)^2 d^2(x_0, y_0) \\ &\quad + (1-s)s (d^2(x_0, y_1) + d^2(x_1, y_0) - d^2(x_0, x_1) - d^2(y_0, y_1)). \end{aligned} \quad (18)$$

and for $x = x_0, y = y_0$ and $x_1 = y_1 = a$ we obtain further

$$\begin{aligned} d^2(\gamma_{x,a}(s), \gamma_{y,a}(t)) &\leq d^2(x, y) + s(s-2) d^2(x, a) + t(t-2) d^2(y, a) \\ &\quad + 2(s+t-st) d(x, a) d(y, a). \end{aligned} \quad (19)$$

Proof. Estimate (16) was proved in [7, Corollary 1.2.5]. Relation (17) can be deduced by applying (D2) twice. Formula (18) follows directly from the previous one by setting $s = t$. It remains to show (19). For $x = x_0, y = y_0$ and $x_1 = y_1 = a$, inequality (17) becomes

$$\begin{aligned} d^2(\gamma_{x,a}(s), \gamma_{y,a}(t)) &\leq (1-s)(1-t) d^2(x, y) + (1-s)(t-s) d^2(x, a) - (1-t)(t-s) d^2(y, a) \\ &= d^2(x, y) - (s+t-st) d^2(x, y) + (1-s)(t-s) d^2(x, a) - (1-t)(t-s) d^2(y, a). \end{aligned}$$

By the triangle inequality we have

$$\begin{aligned} d(x, y) &\geq |d(x, a) - d(y, a)|, \\ d^2(x, y) &\geq d^2(x, a) + d^2(y, a) - 2d(x, a)d(y, a). \end{aligned}$$

Since $s, t \in [0, 1]$, it holds $s+t-st \geq 0$. Replacing $(s+t-st) d^2(x, y)$ we obtain

$$\begin{aligned} d^2(\gamma_{x,a}(s), \gamma_{y,a}(t)) &\leq d^2(x, y) - (s+t-st) (d^2(x, a) + d^2(y, a) - 2d(x, a)d(y, a)) \\ &\quad + (1-s)(t-s) d^2(x, a) - (1-t)(t-s) d^2(y, a). \end{aligned}$$

Resorting yields (19). □

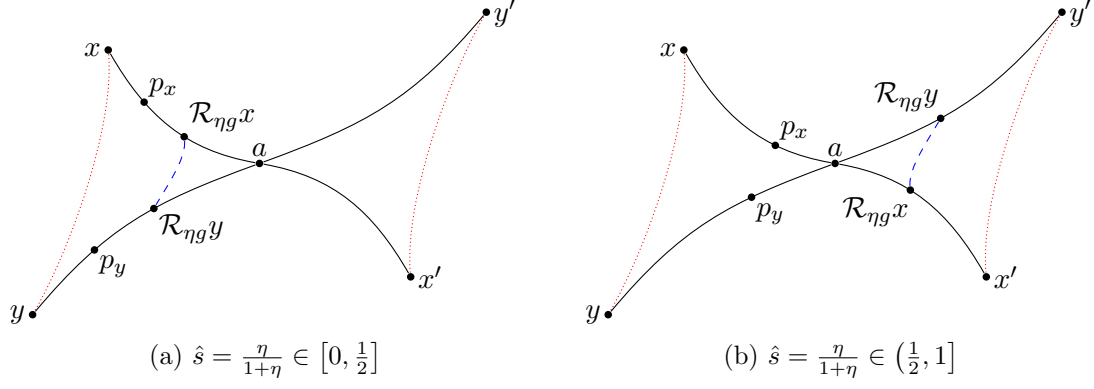


Figure 1. Illustration of the nonexpansiveness of the reflections in Theorem 6.2 for $g(x) = d^2(x, a)$.

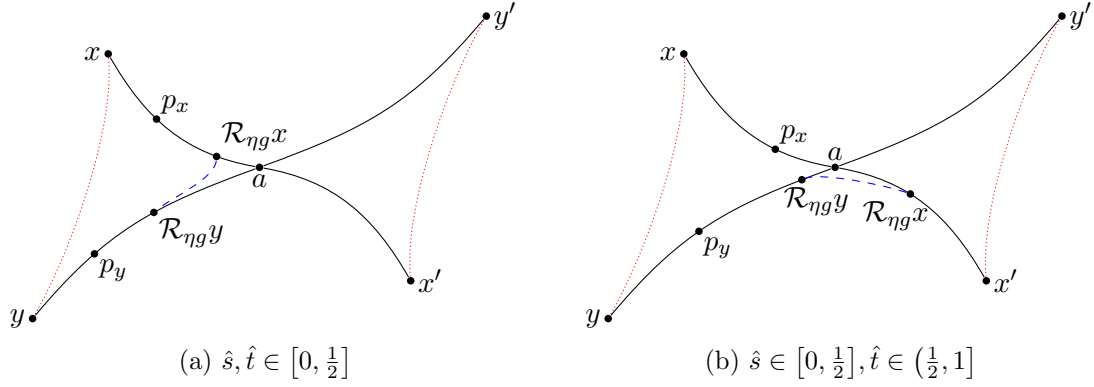


Figure 2. Illustration of the nonexpansiveness of the reflections in Theorem 6.2 for $g(x) = d(x, a)$, where $\hat{s} = \min \left\{ \frac{\eta}{d(x, a)}, 1 \right\}$ and $\hat{t} = \min \left\{ \frac{\eta}{d(y, a)}, 1 \right\}$.

Now we consider the reflections. We will use that for two points $x, a \in \mathcal{H}$ and $p = \gamma_{x,a}(\hat{t})$, $\hat{t} \in [0, 1]$, the reflection of x at p can be written as

$$R_p(x) = \begin{cases} \gamma_{x,a}(2\hat{t}) & \text{if } 2\hat{t} \in [0, 1], \\ \gamma_{x',a}(2 - 2\hat{t}) & \text{if } 2\hat{t} \in (1, 2] \end{cases} \quad (20)$$

with $x' := R_a(x)$.

Theorem 6.2. For an arbitrary fixed $a \in \mathcal{H}$ and $g := \frac{1}{\nu}d^\nu(a, \cdot)$, $\nu \in \{1, 2\}$, the reflection $\mathcal{R}_{\eta g}$, $\eta > 0$, is nonexpansive.

Proof. By (20) we have

$$\mathcal{R}_{\eta g}(x) = R_{p_x}(x) = \begin{cases} \gamma_{x,a}(2\hat{s}) & \text{if } 2\hat{s} \in [0, 1], \\ \gamma_{x',a}(2 - 2\hat{s}) & \text{if } 2\hat{s} \in (1, 2] \end{cases}$$

where $x' := R_a(x)$, and p_x resp. \hat{s} are given by Lemma 5.1 i). Similarly, we get $p_y = \text{prox}_{\eta g}(y) = \gamma_{y,a}(\hat{t})$, where $\hat{t} = \hat{s}$ if $\nu = 2$ and $\hat{t} = \min \left\{ \frac{\eta}{d(y, p)}, 1 \right\}$ if $\nu = 1$.

1. Let $\nu = 2$. We distinguish two cases.

1.1. If $2\hat{s} \in [0, 1]$, cf. Fig. 1 (a), we obtain with $s := 2\hat{s}$ by the joint convexity of the distance function

$$d(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) = d(\gamma_{\widehat{x}, a}(s), \gamma_{\widehat{y}, a}(s)) \leq (1-s)d(x, y) \leq d(x, y).$$

1.2. In the case $2\hat{s} \in (1, 2]$, see Fig. 1 (b), we verify with $s := 2 - 2\hat{s} \in (0, 1]$ by the joint convexity of the distance function and since the reflection is an isometry that

$$d(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) = d(\gamma_{\widehat{x}, a}(s), \gamma_{\widehat{y}, a}(s)) \leq (1-s)d(R_a(x), R_a(y)) \leq d(x, y).$$

2. Let $\nu = 1$. Without loss of generality we assume that $d(x, a) \leq d(y, a)$ so that $\frac{\eta}{d(y, a)} \leq \frac{\eta}{d(x, a)}$. We distinguish three cases.

2.1. In the case $d(x, a) \leq d(y, a) \leq \eta$ we obtain $\hat{s} = \hat{t} = 1$ and consequently we set $s := 2 - 2\hat{s} = 0$ and $t := 2 - 2\hat{t} = 0$. Since the reflection is an isometry this implies

$$d(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) = d(\gamma_{\widehat{x}, a}(0), \gamma_{\widehat{y}, a}(0)) = d(R_a(x), R_a(y)) = d(x, y).$$

2.2. In the case $\eta \leq d(x, a) \leq d(y, a)$ we have $\hat{s} = \frac{\eta}{d(x, a)}$ and $\hat{t} = \frac{\eta}{d(y, a)}$. We distinguish three cases.

i) If $\hat{t}, \hat{s} \leq \frac{1}{2}$, cf. Fig. 2 (a), we obtain with $s := 2\hat{s}$ and $t := 2\hat{t}$ by (19) the estimate

$$\begin{aligned} d^2(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) &= d^2(\gamma_{\widehat{x}, a}(s), \gamma_{\widehat{y}, a}(t)) \\ &\leq d^2(x, y) + s(s-2)d^2(x, a) + t(t-2)d^2(y, a) \\ &\quad + 2(s+t-st)d(x, a)d(y, a). \end{aligned}$$

Plugging in $s = \frac{2\eta}{d(x, a)}$ and $t = \frac{2\eta}{d(y, a)}$ we see that the expressions containing s and t vanish so that the reflection is non-expansive.

ii) If $\hat{t} \leq \frac{1}{2} \leq \hat{s}$, cf. Fig. 2 (b), we get with $s := 2 - 2\hat{s}$ and $t := 2\hat{t}$ that

$$d(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) = d(\gamma_{\widehat{x}, a}(s), \gamma_{\widehat{y}, a}(t)) = d(\tilde{x}, \tilde{y})$$

with $\tilde{x} = \gamma_{\widehat{x}, a}(s)$ and $\tilde{y} = \gamma_{\widehat{y}, a}(t)$. By the assumption we have $d(a, \tilde{x}) = 2\eta - d(x, a)$ and $d(y, a) = 2\eta + d(a, \tilde{y})$. Further, it follows by the triangle inequality

$$\begin{aligned} d(\tilde{y}, \tilde{x}) &\leq d(a, \tilde{x}) + d(a, \tilde{y}) = 2\eta - d(x, a) + d(a, \tilde{y}), \\ d(x, y) &\geq d(y, a) - d(x, a) = 2\eta + d(a, \tilde{y}) - d(x, a). \end{aligned}$$

Hence we conclude $d(\tilde{y}, \tilde{x}) \leq d(x, y)$.

iii) If $\hat{t}, \hat{s} \geq \frac{1}{2}$ we set $s := 2 - 2\hat{s}$ and $t := 2 - 2\hat{t}$ and obtain by (19) and the isometry property of the reflection

$$\begin{aligned} d^2(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) &= d^2(\gamma_{\widehat{x}, a}(s), \gamma_{\widehat{y}, a}(t)) \\ &\leq d^2(x, y) + s(s-2)d^2(x, a) + t(t-2)d^2(y, a) \\ &\quad + 2(s+t-st)d(x, a)d(y, a). \end{aligned}$$

Plugging in the expressions for s and t and noting that the above expression is symmetric with respect to (s, t) and $(2-s, 2-t)$ we obtain the assertion as in i).

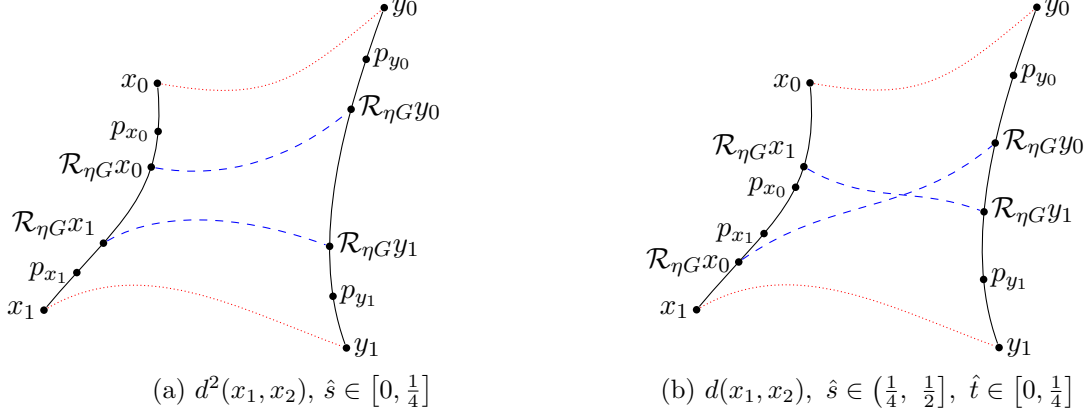


Figure 3. Illustration of the nonexpansiveness of the reflections in Theorem 6.3, where $p_x = (p_{x_0}, p_{x_1})$ and $p_y = (p_{y_0}, p_{y_1})$, see Lemma 5.1. (a) case 1., i.e., $\nu = 2$ and $\hat{s} \in [0, \frac{1}{4}]$, (b) case 2.2., i.e., $\nu = 1$, $\hat{s} = \frac{1}{2}$, and $\hat{t} \in [0, \frac{1}{4}]$.

2.3. In the case $d(x, a) \leq \eta \leq d(y, a)$ we have $\hat{s} = 1$ and $\hat{t} = \frac{\eta}{d(y, a)}$. Again, we distinguish two cases:

i) If $\hat{t} \leq \frac{1}{2}$ we use $t := 2\hat{t}$ and obtain

$$d(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) = d(\gamma_{\widehat{x', a}}(0), \gamma_{\widehat{y, a}}(t)) = d(x', \tilde{y})$$

with $\tilde{y} := \gamma_{\widehat{y, a}}(t)$. By assumption we have $2d(x, a) \leq 2\eta$ so that

$$d(a, y) \geq 2d(a, x) + d(a, \tilde{y}).$$

By the triangle inequality it follows

$$\begin{aligned} d(x', \tilde{y}) &\leq d(x', a) + d(a, \tilde{y}) = d(x, a) + d(a, \tilde{y}), \\ d(x, y) &\geq d(a, y) - d(a, x) \geq d(a, x) + d(a, \tilde{y}) \end{aligned}$$

and consequently $d(x', \tilde{y}) \leq d(x, y)$.

ii) If $\hat{t} > \frac{1}{2}$ we put $t = 2 - 2\hat{t}$ and get by (19) and since the reflection is an isometry

$$\begin{aligned} d^2(\mathcal{R}_{\eta g}(x), \mathcal{R}_{\eta g}(y)) &= d^2(\gamma_{\widehat{x', a}}(0), \gamma_{\widehat{y, a}}(t)) \\ &\leq d^2(x, y) + 2td(x, a)d(y, a) + t(t - 2)d^2(y, a) \\ &= d^2(x, y) + 2td(y, a)(d(x, a) - \eta) \leq d^2(x, y). \end{aligned}$$

This finishes the proof. □

Recall that the distance $d_{\mathcal{H}^2}$ on $\mathcal{H}^2 = \mathcal{H} \times \mathcal{H}$ is defined by

$$d_{\mathcal{H}^2}((x_0, x_1), (y_0, y_1)) := \sqrt{d^2(x_0, y_0) + d^2(x_1, y_1)}.$$

Theorem 6.3. For $G := d^\nu(\cdot, \cdot)$, $\nu \in \{1, 2\}$, the reflection $\mathcal{R}_{\eta G}$, $\eta > 0$, is nonexpansive.

Proof. By (20) and since $s := 2\hat{s} \in [0, 1]$ we have

$$\mathcal{R}_{\eta G}(x) = R_{p_x}(x) = (\gamma_{x_0, x_1}(s), \gamma_{x_1, x_0}(s)), \quad (21)$$

where p_x is given by Lemma 5.1 ii). Similarly, we conclude for $y := (y_0, y_1)$ with $\hat{t} = \hat{s}$ if $\nu = 2$ and $\hat{t} = \min\{\frac{\eta}{d(y_0, y_1)}, \frac{1}{2}\}$ if $\nu = 1$.

1. Let $\nu = 2$, cf. Fig. 3 (a). We obtain

$$d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) = d^2(\gamma_{x_0, x_1}(s), \gamma_{y_0, y_1}(s)) + d^2(\gamma_{x_1, x_0}(s), \gamma_{y_1, y_0}(s))$$

and applying (18) we further have

$$\begin{aligned} d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) &\leq (s^2 + (1-s)^2)(d^2(x_0, y_0) + d^2(x_1, y_1)) \\ &\quad + 2(1-s)s(d^2(x_0, y_1) + d^2(x_1, y_0) - d^2(x_0, x_1) - d^2(y_0, y_1)) \\ &= d^2(x_0, y_0) + d^2(x_1, y_1) + 2s(s-1)(d^2(x_0, y_0) + d^2(x_1, y_1)) \\ &\quad - 2s(s-1)(d^2(x_0, y_1) + d^2(x_1, y_0) - d^2(x_0, x_1) - d^2(y_0, y_1)) \\ &= d_{\mathcal{H}^2}^2(x, y) - 2s(1-s) \\ &\quad \times (d^2(x_0, y_0) + d^2(x_1, y_1) + d^2(x_0, x_1) + d^2(y_0, y_1) - d^2(x_0, y_1) - d^2(x_1, y_0)). \end{aligned}$$

By (16) we know that the last factor is non-negative. Since $s \in [0, 1]$ we conclude $2s(1-s) \geq 0$ so that

$$d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) \leq d_{\mathcal{H}^2}^2(x, y).$$

2. Let $\nu = 1$. Without loss of generality we assume that $d(x_0, x_1) \leq d(y_0, y_1)$ and consequently $\frac{\eta}{d(y_0, y_1)} \leq \frac{\eta}{d(x_0, x_1)}$. We distinguish three cases.

2.1. If $d(x_0, x_1) \leq d(y_0, y_1) \leq 2\eta$ we have $s = 1$ so that by (21) it holds

$$d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) = d^2(x_1, y_1) + d^2(x_0, y_0) = d_{\mathcal{H}^2}^2(x, y).$$

2.2. If $2\eta \leq d(x_0, x_1) \leq d(y_0, y_1)$, cf. Fig. 3 (b), we use (17) to obtain

$$\begin{aligned} d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) &= d^2(\gamma_{x_0, x_1}(s), \gamma_{y_0, y_1}(t)) + d^2(\gamma_{x_1, x_0}(s), \gamma_{y_1, y_2}(1-t)) \\ &\leq (st + (1-s)(1-t))(d^2(x_0, y_0) + d^2(x_1, y_1)) \\ &\quad + ((1-s)t + s(1-t))(d^2(x_0, y_1) + d^2(x_1, y_0)) \\ &\quad - 2(s-s^2)d^2(x_0, x_1) - 2(t-t^2)d^2(y_0, y_1) \\ &= d^2(x_0, y_0) + d^2(x_1, y_1) \\ &\quad + (t+s-2st)(d^2(x_0, y_1) + d^2(x_1, y_0) - d^2(x_0, y_0) - d^2(x_1, y_1)) \\ &\quad - 2(s-s^2)d^2(x_0, x_1) - 2(t-t^2)d^2(y_0, y_1). \end{aligned}$$

Since $t, s \in (0, 1]$ we get $s + t - 2st > 0$ and we can use (8) to estimate

$$d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) \leq d^2(x_0, y_0) + d^2(x_1, y_1) + (s + t - 2st)2d(x_0, x_1)d(y_0, y_1) - 2(s - s^2)d^2(x_0, x_1) - 2(t - t^2)d^2(y_0, y_1). \quad (22)$$

Plugging in the definitions of s, t we conclude

$$\begin{aligned} d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) &\leq d^2(x_0, y_0) + d^2(x_1, y_1) \\ &\quad + 2\left(\frac{2\eta}{d(x_0, x_1)} + \frac{2\eta}{d(y_0, y_1)} - \frac{8\eta^2}{d(x_0, x_1)d(y_0, y_1)}\right)d(x_0, x_1)d(y_0, y_1) \\ &\quad - 2\left(\frac{2\eta}{d(x_0, x_1)} - \frac{4\eta^2}{d^2(x_0, x_1)}\right)d^2(x_0, x_1) - 2\left(\frac{2\eta}{d(y_0, y_1)} - \frac{4\eta^2}{d^2(y_0, y_1)}\right)d^2(y_0, y_1) \\ &= d_{\mathcal{H}^2}^2((x_0, x_1), (y_0, y_1)). \end{aligned}$$

2.3. For $d(x_0, x_1) \leq 2\eta \leq d(y_0, y_1)$ we have $s = 1$ and $t = \frac{2\eta}{d(y_0, y_1)}$. Substituting these values into (22) we obtain

$$\begin{aligned} d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) &\leq d^2(x_0, y_0) + d^2(x_1, y_1) \\ &\quad + \left(1 - \frac{2\eta}{d(y_0, y_1)}\right)2d(x_0, x_1)d(y_0, y_1) - 4\eta d(y_0, y_1) + 8\eta^2 \end{aligned}$$

and using $d(x_0, x_1) \leq 2\eta$ we obtain further

$$d_{\mathcal{H}^2}^2(\mathcal{R}_{\eta G}(x), \mathcal{R}_{\eta G}(y)) \leq d^2(x_0, y_0) + d^2(x_1, y_1) = d_{\mathcal{H}^2}^2(x, y).$$

This finishes the proof. \square

7. Reflections on Convex Sets

Next we deal with the reflection operator $\mathcal{R}_{\mathcal{D}}$, i.e., the reflection operator which corresponds to the orthogonal projection operator onto the convex set \mathcal{D} . Unfortunately, in symmetric Hadamard manifolds, reflections corresponding to orthogonal projections onto convex sets are in general not nonexpansive. Counterexamples can be found in [19, 32]. Unfortunately this is also true for our special set \mathcal{D} as the following example with symmetric positive definite 2×2 matrices $\mathcal{P}(2)$ shows. For the manifold $\mathcal{P}(n)$ of symmetric positive definite matrices see Appendix B.

Example 7.1. Let $\mathcal{D} = \{(x, x, x) : x \in \mathcal{P}(2)\}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{P}(2)^3$ be given by

$$\begin{aligned} \mathbf{x} &= \left(\begin{pmatrix} 20.9943 & 3.3101 \\ 3.3101 & 6.8906 \end{pmatrix}, \begin{pmatrix} 17.2428 & 4.3111 \\ 4.3111 & 9.9950 \end{pmatrix}, \begin{pmatrix} 19.4800 & 19.8697 \\ 19.8697 & 21.2513 \end{pmatrix} \right), \\ \mathbf{y} &= \left(\begin{pmatrix} 7.5521 & 6.0509 \\ 6.0509 & 19.8961 \end{pmatrix}, \begin{pmatrix} 6.4261 & 5.7573 \\ 5.7573 & 15.2775 \end{pmatrix}, \begin{pmatrix} 12.4792 & 12.9202 \\ 12.9202 & 13.8620 \end{pmatrix} \right). \end{aligned}$$

The distance between \mathbf{x} and \mathbf{y} is $d_{\mathcal{P}(2)}(\mathbf{x}, \mathbf{y}) \approx 2.2856$.

The projection onto D can be calculated using the gradient descent method from [2]. We obtain

$$\begin{aligned}\text{prox}_{\iota_D}(\mathbf{x}) &\approx (1, 1, 1)^T \otimes \begin{pmatrix} 13.8254 & 8.7522 \\ 8.7522 & 10.8436 \end{pmatrix}, \\ \text{prox}_{\iota_D}(\mathbf{y}) &\approx (1, 1, 1)^T \otimes \begin{pmatrix} 8.3908 & 8.2797 \\ 8.2797 & 12.4013 \end{pmatrix},\end{aligned}$$

where \otimes denotes the Kronecker product. For the distance between them we obtain

$$d_{\mathcal{P}(2)}(\mathcal{R}_{\iota_D}\mathbf{x}, \mathcal{R}_{\iota_D}\mathbf{y}) \approx 2.7707 > d_{\mathcal{P}(2)}(\mathbf{x}, \mathbf{y}),$$

i.e., the reflection is not nonexpansive. The computations were done with machine precision in MATLAB and the results are rounded to four digits.

The situation changes if we consider manifolds with constant curvature κ . We do not restrict ourselves to Hadamard manifolds now, but deal instead with the model spaces for constant curvature \mathcal{M}_κ^d which are defined as follows: For $\kappa > 0$ the model space \mathcal{M}_κ^d is obtained from the d -dimensional sphere by multiplying the distance with $\frac{1}{\sqrt{\kappa}}$; for $\kappa < 0$ we get \mathcal{M}_κ^d from the d -dimensional hyperbolic plane by multiplying the distance with $\frac{1}{\sqrt{-\kappa}}$; finally \mathcal{M}_0^d is the d -dimensional Euclidean space. The model spaces inherit their geometrical properties from the three Riemannian manifolds that define them. Thus, if $\kappa < 0$, then \mathcal{M}_κ^d is uniquely geodesic, balls are convex and we have a counterpart for the hyperbolic law of cosine. By r_κ we denote the convex radius of the model spaces \mathcal{M}_κ^d which is the supremum of radii of balls, which are convex in \mathcal{M}_κ^d , i.e.,

$$r_\kappa := \begin{cases} \infty, & \kappa \leq 0, \\ \frac{\pi}{\sqrt{\kappa}}, & \text{otherwise.} \end{cases}$$

To show, that reflections at convex sets in manifolds with constant curvature are nonexpansive, we need the following properties of projections onto convex sets.

Proposition 7.2. [19, Proposition II.2.4, Exercise II.2.6(1)]

Let X be a complete $\text{CAT}(\kappa)$, $\kappa \in \mathbb{R}$, space, $V = \{x : d(x, V) \leq r_\kappa/2\}$, $x \in V$, and $C \subset X$ is closed and convex. Then the following statements hold

1. The metric projection $\Pi_C(x)$ of x onto C is a singleton.
2. If $y \in [x, \Pi_C(x)]$, then $\Pi_C(x) = \Pi_C(y)$.
3. If $x \notin C$, $y \in C$, and $y \neq \Pi_C(x)$, then $\angle_{\Pi_C(x)}(x, y) \geq \frac{\pi}{2}$, where $\angle_{\Pi_C(x)}(x, y)$ denotes the angle at $\Pi_C(x)$ between the geodesics $\gamma_{\Pi_C(x), x}$ and $\gamma_{\Pi_C(x), y}$.

Theorem 7.3. Let $k \in \mathbb{R}$ and $d \in \mathbb{N}$. Suppose that \mathcal{C} is a nonempty closed and convex subset of \mathcal{M}_κ^d . Let $x, y \in \mathcal{M}_\kappa^d$ such that $\text{dist}(x, \mathcal{C}), \text{dist}(y, \mathcal{C}) \leq r_\kappa$ be given. Then

$$d(\mathcal{R}_{\iota_{\mathcal{C}}}x, \mathcal{R}_{\iota_{\mathcal{C}}}y) \leq d(x, y).$$

Proof. The case $\kappa < 0$ was proved in [32]. For $\kappa = 0$ the assumption follows from the Hilbert space setting. We adapt the proof from [32] to show the case $\kappa > 0$. Due to the structure of the model spaces it is sufficient to show the nonexpansiveness for $\kappa = 1$.

For simplicity, we denote $c_x = \Pi_C(x)$, $c_y = \Pi_C(y)$, $x' = R_C(x)$, $y' = R_C(y)$, $\alpha = \angle_{c_x}(x, c_y)$, and $\alpha' = \angle_{c_x}(y, c_x)$. Notice, that by Proposition 7.2, $\alpha, \alpha' \geq \frac{\pi}{2}$, and $d(x, y) \leq \pi, \forall y, x \in \mathbb{S}^n$. Consider the geodesic triangles $\triangle x c_x c_y$ and $\triangle x' c_x c_y$. By the spherical law of cosines we have that

$$\cos(d(x, c_y)) = \cos(d(x, c_x)) \cos(d(c_x, c_y)) + \sin(d(x, c_x)) \sin(d(c_x, c_y)) \cos(\alpha)$$

and

$$\cos(d(x', c_y)) = \cos(d(x', c_x)) \cos(d(c_x, c_y)) + \sin(d(x', c_x)) \sin(d(c_x, c_y)) \cos(\pi - \alpha).$$

Since $d(x, c_x) = d(x', c_x)$ and $\cos(\pi - \alpha) \geq 0$ we get

$$\cos(d(x, c_y)) \leq \cos(d(x', c_y)). \quad (23)$$

Similarly we get

$$\cos(d(y, c_x)) \leq \cos(d(y', c_x)). \quad (24)$$

Consider now the geodesic triangles $\triangle x' c_x c_y$ and $\triangle x' c_x c_y$ and denote $\beta = \angle_{c_x}(x, y)$. Applying again the spherical law of cosines we obtain that

$$\cos(d(x, y)) = \cos(d(x, c_x)) \cos(d(c_x, y)) + \sin(d(x, c_x)) \sin(d(c_x, y)) \cos(\beta)$$

and

$$\cos(d(x', y)) = \cos(d(x', c_x)) \cos(d(c_x, y)) + \sin(d(x', c_x)) \sin(d(c_x, y)) \cos(\pi - \beta).$$

Since $d(x, c_x) = d(x', c_x)$ and $\cos(\pi - \beta) = -\cos \beta$, we get by adding

$$\cos(d(x, y)) + \cos(d(x', y)) = 2 \cos(d(x, c_x)) \cos(d(c_x, y)) \quad (25)$$

and similar

$$\cos(d(x', y')) + \cos(d(x, y')) = 2 \cos(d(x, c_x)) \cos(d(c_x, y')), \quad (26)$$

$$\cos(d(x, y)) + \cos(d(x, y')) = 2 \cos(d(x, c_y)) \cos(d(c_y, y)), \quad (27)$$

$$\cos(d(x', y')) + \cos(d(x', y)) = 2 \cos(d(x', c_y)) \cos(d(c_y, y)). \quad (28)$$

Suppose now that $d(x', y') > d(x, y)$. From (25), (24), and (26) we obtain

$$\begin{aligned} \cos(d(x', y')) + \cos(d(x, y')) &= 2 \cos(d(x, c_x)) \cos(d(c_x, y')) \\ &\geq 2 \cos(d(x, c_x)) \cos(d(c_x, y)) \\ &= \cos(d(x, y)) + \cos(d(x', y)) \\ &> \cos(d(x', y')) + \cos(d(x', y)) \end{aligned}$$

which implies $d(x, y') < d(x', y)$. Using (27), (23), and (28) we get

$$\begin{aligned} \cos(d(x, y)) + \cos(d(x, y')) &= 2 \cos(d(x, c_y)) \cos(d(c_y, y)) \\ &\leq 2 \cos(d(x', c_y)) \cos(d(c_y, y)) \\ &= \cos(d(x', y')) + \cos(d(x', y)) \\ &< \cos(d(x, y)) + \cos(d(x', y)) \end{aligned}$$

which is a contradiction to $d(x, y') < d(x', y)$, thus the result follows. \square

8. Numerical Examples

In this section, we demonstrate the performance of the Parallel DR Algorithm 4 (PDRA) by numerical examples. We chose a constant step size $\lambda_r := \lambda \in (0, 1)$ in the reflections for all $r \in \mathbb{N}$, where the specific value of λ is stated in each experiment.

We compare our algorithm with two other approaches to minimize the functional \mathcal{E} in (12), namely the Cyclic Proximal Point Algorithm (CPPA) [8, 79] and the Half-Quadratic Minimization Algorithm in its multiplicative form (HQMA) [14]. Let us briefly recall these algorithms.

CPPA. Based on the splitting (13) of our functional (12) the CPPA computes the iterates

$$u^{(r+1)} = \text{prox}_{\eta_r \varphi_5} \circ \dots \circ \text{prox}_{\eta_r \varphi_1}(u^{(r)}).$$

The algorithm converges for any starting point $u^{(0)}$ to a minimizer of the functional (12) supposed that

$$\sum_{r \in \mathbb{N}} \eta_r = +\infty, \quad \sum_{r \in \mathbb{N}} \eta_r^2 < +\infty.$$

Therefore the sequence $\{\eta_r\}_r$ must decrease. In our computations we chose $\eta_r := \frac{\eta}{r+1}$, where the specific value of η is addressed in the experiments.

HQMA. The HQMA cannot be directly applied to \mathcal{E} , but to its differentiable substitute

$$\mathcal{E}_\delta(u) := \frac{1}{2} \sum_{(i,j) \in \mathcal{V}} d(f_{i,j}, u_{i,j})^2 + \alpha \left(\sum_{(i,j) \in \mathcal{G}} \varphi(d(u_{i,j}, u_{i+1,j})) + \sum_{(i,j) \in \mathcal{G}} \varphi(d(u_{i,j}, u_{i,j+1})) \right),$$

where $\varphi(t) := \sqrt{t^2 + \delta^2}$, $\delta > 0$. For small δ the functional \mathcal{E}_δ approximates \mathcal{E} . Other choices of φ are possible, see, e.g., [14, 60]. The functional can be rewritten as

$$\begin{aligned} \mathcal{E}_\delta(u) &= \min_{\mathbf{v}, \mathbf{w} \in \mathbb{R}^{N,M}} \left\{ \frac{1}{2} \sum_{(i,j) \in \mathcal{V}} d(f_{i,j}, u_{i,j})^2 \right. \\ &\quad \left. + \alpha \left(\sum_{(i,j) \in \mathcal{G}} v_{i,j} d^2(u_{i,j}, u_{i+1,j}) + \psi(v_{i,j}) + w_{i,j} d^2(u_{i,j}, u_{i,j+1}) + \psi(w_{i,j}) \right) \right\} \\ &= \min_{\mathbf{v}, \mathbf{w} \in \mathbb{R}^{N,M}} E(u, \mathbf{v}, \mathbf{w}), \end{aligned}$$

where $\psi(s) := \min_{s \in \mathbb{R}} \{t^2 s - \varphi(t)\}$, see [14]. Then an alternating algorithm is used to minimize the right-hand side, i.e., the HQMA computes

$$(\mathbf{v}^{(r)}, \mathbf{w}^{(r)}) = \arg \min_{\mathbf{v}, \mathbf{w} \in \mathbb{R}^{N,M}} E(u^{(r)}, \mathbf{v}, \mathbf{w}), \quad (29)$$

$$u^{(r+1)} = \arg \min_{u \in \mathcal{H}^{N,M}} E(u, \mathbf{v}^{(r)}, \mathbf{w}^{(r)}). \quad (30)$$

For (29) there exists an analytic expression, while the minimization in (30) is done with a Riemannian Newton method [1]. The whole HQMA can be considered as a quasi-Newton algorithm to minimize \mathcal{E}_δ . The convergence of $\{u^{(r)}\}_r$ to a minimizer of \mathcal{E}_δ is ensured by [14, Theorem 3.4].

In order to compare the algorithms, we implemented them in a common framework in MATLAB. They rely on the same manifold functions, like the distance function, the proximal mappings, and the exponential and logarithmic map. These basic ingredients as well as the Riemannian Newton step in the HQMA, especially the Hessian needed therein, were implemented in C++ and imported into MATLAB using `mex`-interfaces while for all algorithms the main iteration was implemented in MATLAB for convenience. The implementation is based on the `Eigen` library¹ 3.2.4. All tests were run on a MacBook Pro running Mac OS X 10.11.1, Core i5, 2.6 GHz, with 8 GB RAM using MATLAB 2015b and the clang-700.1.76 compiler.

As *initialization* of the iteration we use $u^{(0)} = f$ if all pixels of the initial image f are known and a nearest neighbor initialization for the missing pixels otherwise. As *stopping criterion* we employ $\epsilon^{(r)} := d(x^{(r)}, x^{(r-1)}) < \epsilon$ often in combination with a maximal number of permitted iteration steps r_{\max} .

We consider three manifolds:

1. The space of univariate Gaussian probability distributions with the Fisher metric: As in [4], we associate a univariate Gaussian probability distribution to each image pixel in Subsection 8.1. Then we denoise this image by our approach.
2. Symmetric positive definite 2×2 matrices of determinant 1 with the affine invariant Riemannian metric: Following an idea of [23], we model structure tensors as elements in this space and denoise the tensor images in Subsection 8.2.
3. Symmetric positive definite 3×3 matrices with the affine invariant Riemannian metric: We are interested in denoising diffusion tensors in DT-MRI in Subsection 8.3. While the first two experiments deal only with the denoising of images, we further perform a combined inpainting and denoising approach in Subsection 8.4.

The first two settings are manifolds with constant curvature $-\frac{1}{2}$. Indeed, Appendix C shows that there are isomorphisms between these spaces and the hyperbolic manifold \mathbb{H}_M . Our implementations use these isomorphisms to work finally on the hyperbolic manifold \mathbb{H}_M . The symmetric positive definite matrices in the third setting form a symmetric Hadamard manifold, but its curvature is not constant. Although the reflection \mathcal{R}_{ι_D} is in general not nonexpansive which is required in the convergence proof of Theorem 4.3, we observed convergence in all our numerical examples.

8.1. Univariate Gaussian Distributions

In this section, we deal with a series of m similar gray-value images $(g_{i,j}^k)_{(i,j) \in \mathcal{G}}$, $k = 1, \dots, m$. We assume that the gray-values $g_{i,j}^k$, $k = 1, \dots, m$, of each pixel $(i, j) \in \mathcal{G}$ are realizations of a univariate Gaussian random variable with distribution $N(\mu_{i,j}, \sigma_{i,j})$. We estimate the corresponding mean and the standard deviation by the maximum likelihood estimators

$$\mu_{i,j} = \frac{1}{m} \sum_{k=1}^m g_{i,j}^k, \quad \sigma_{i,j} = \sqrt{\frac{1}{m} \sum_{k=1}^m (g_{i,j}^k - \mu_{i,j})^2}. \quad (31)$$

¹open source, available at <http://eigen.tuxfamily.org>

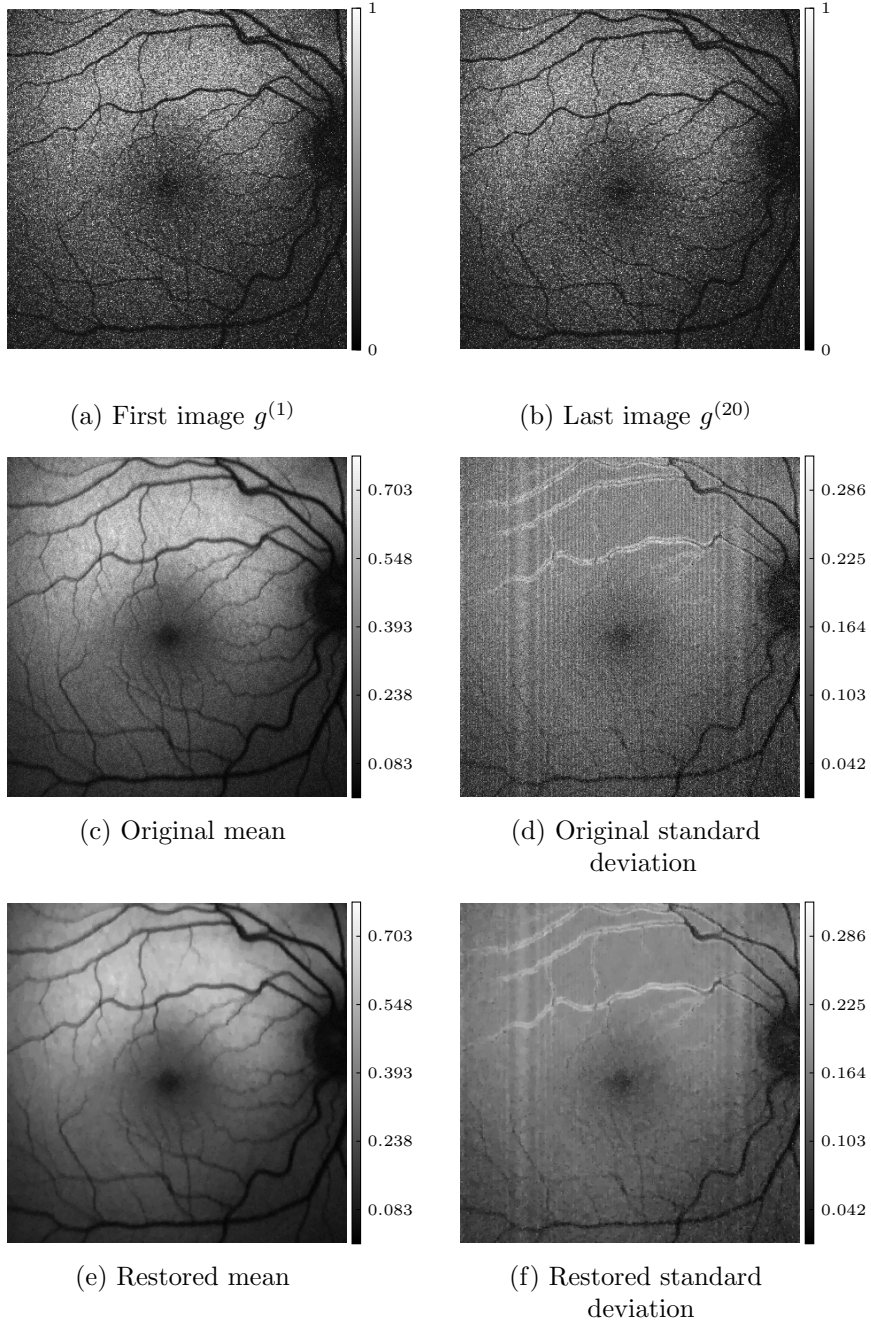


Figure 4. Denoising of the retina data using model (12) with $\alpha = 0.2$ and the PDRA ($\eta = \frac{1}{2}$, $\lambda = \frac{9}{10}$). The restored image keeps the main features, e.g. veins in the mean and their movement in the variance in an area around them.

η	CPPA in sec.	PDRA		
		$\lambda = 0.5$	$\lambda = 0.9$	$\lambda = 0.95$
0.05	56.85	129.26	65.21	59.84
0.1	56.54	59.21	34.32	36.67
0.5	65.17	57.41	42.06	46.07
1	57.14	93.75	63.58	58.66

Table 1. Computation times for the CPPA and PDRA for different values of η and λ in seconds.

η	CPPA 184.3643+...	PDRA		
		$\lambda = 0.5$	$\lambda = 0.9$	$\lambda = 0.95$
0.05	44.80	1.021×10^{-5}	1.180×10^{-5}	1.627×10^{-5}
0.1	10.65	2.514×10^{-5}	2.969×10^{-5}	3.429×10^{-5}
0.5	1.055×10^{-2}	5.082×10^{-4}	2.785×10^{-4}	2.256×10^{-4}
1	1.953×10^{-2}	8.189×10^{-4}	5.027×10^{-4}	4.992×10^{-4}

Table 2. Distance of the computed minimum to the assumed one of 184.3643.

We consider images $f: \mathcal{G} \rightarrow \mathcal{N}$ mapping into the Riemannian manifold \mathcal{N} of univariate non-degenerate Gaussian probability distributions, parameterized by the mean and the standard deviation, with the Fisher metric. For the definition of the Fisher metric as well as its relation to the hyperbolic manifold, see Appendix C.

In our numerical examples we use a sequence of $m = 20$ images $(g_{i,j}^k)_{i,j=1}^{384}$ of size 384×384 taken from the same retina by a CCD (coupled charged-device) camera, cf. [4, Fig. 13], over a very short time frame, each with a very short exposure time. Hence, we have noisy images that are also affected by the movement of the eye. In the top row of Fig. 4 we see the first and the last image of this sequence. From these images, we obtain the —still noisy— image $f: \mathcal{G} \rightarrow \mathcal{N}$ with $(f_{i,j})_{i,j=1}^{384} = ((\mu_{i,j}, \sigma_{i,j}))_{i,j=1}^{384}$ depicted in the middle row of Fig. 4 by (31). We denoise this image by minimizing the functional (12) with $\alpha = 0.2$ by the PDRA ($\eta = \frac{1}{2}$, $\lambda = \frac{9}{10}$). The result is shown in the bottom row of Fig. 4.

Next we compare the performance of the PDRA with the CPPA for different values η of the proximal mapping and various parameters λ of the reflection, both applied to a subset of 50×50 pixels of the Retina dataset. The stopping criterion is $r_{\max} = 1500$ iterations or $\epsilon = 10^{-6}$.

Table 1 records the computational time of both algorithms. The CPPA requires always 1500 iterations which takes roughly a minute. It appears to be robust with respect to the choice of η . The PDRA on the other hand depends on the chosen value of η , e.g., for $\lambda = 0.5$ the computation time for a small η is 129 seconds, while its minimal computation time is just 34 seconds, for $\lambda = 0.9$ and $\eta = 0.1$. For the latter parameters, the iteration stops after 278 iterations with an $\epsilon^{(278)} < 10^{-6}$. The runtime per iteration is longer than for the CPPA due to the Karcher mean computation in each iteration which is implemented using the gradient descent method of [2].

While the computation time does not change much for the CPPA with different η , the resulting values of the functional \mathcal{E} in (12) do significantly. We compare the same values for

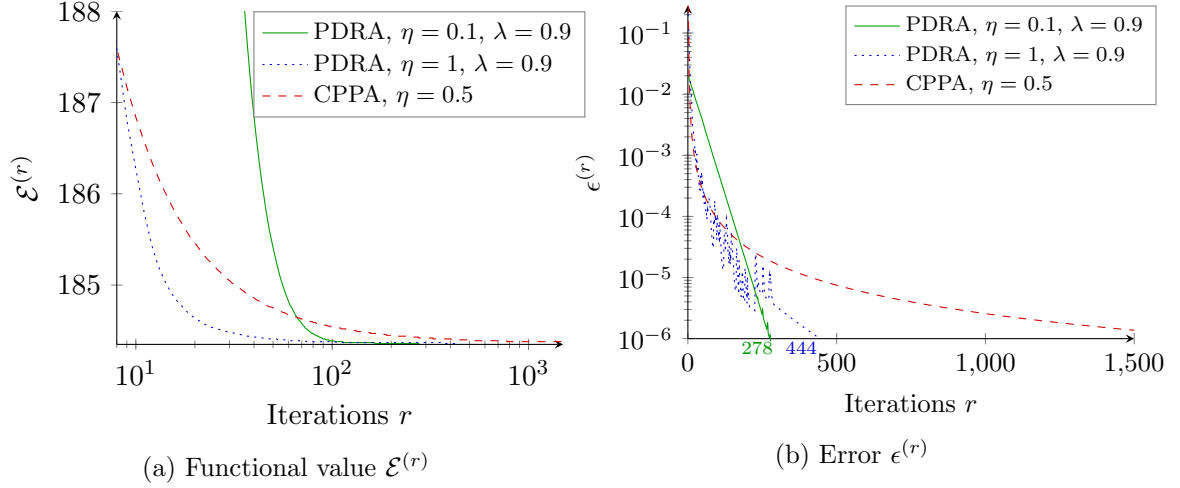


Figure 5. Comparison of (a) the functional value $\mathcal{E}^{(r)}$ and (b) the error $\epsilon^{(r)}$ for the CPPA and the PDRA for different parameters. The PDRA converges faster both with respect to the functional value and the error.

η and λ as for the time measurements in Table 2. To this end, we estimated the minimum of the functional as 184.3643 by performing the PDRA with 3000 iterations. We use this value to compare the results obtained for different parameters with this estimated minimum. Though the fastest PDRA ($\lambda = 0.9$, $\eta = 0.1$) does not yield the lowest value of \mathcal{E} , the slowest one ($\lambda = 0.5$, $\eta = 0.05$) does. Looking at the values for the CPPA, they depend heavily on the parameter η and are further away from the minimum than all PDRA tests.

Fig. 5(a) left shows the development of $\mathcal{E}^{(r)} := \mathcal{E}(u^{(r)})$ with respect to the number of iterations. The CPPA converges quite slowly. Both plotted PDRA tests, the fastest (solid green) and the one starting with the steepest descent (dashed blue) from Table 2, decay faster. Due to the fast decay of $\mathcal{E}^{(r)}$ at the beginning, the iteration number is shown in a logarithmic scale. Furthermore, also the errors $\epsilon^{(r)} := d(x^{(r)}, x^{(r-1)})$, which are plotted in a logarithmic scale in Fig. 5(b), decay much faster for the PDRA than for the CPPA. Nevertheless, for $\eta = 1$, the decay is not monotone at the beginning of the algorithm.

In summary, the PDRA performs better than the CPPA with respect to runtime and minimal functional value. It is nearly independent of the choice of η when looking at the functional values in contrast to the CPPA. However, both η and λ have an influence on the runtime of the algorithm.

8.2. Scaled Structure Tensor

The structure tensor of Förstner and Gülch [34], see also [77], can be used to determine edges and corners in images. For each pixel $x \in \mathcal{G}$ of an image $f: \mathcal{G} \rightarrow \mathbb{R}$ it is defined as the 2×2 matrix

$$\mathcal{J}_\rho(x) := (G_\rho * \nabla f_\sigma \nabla f_\sigma^T)(x) = \sum_{y \in \mathcal{G}} G_\rho(x - y) \nabla f_\sigma(y) \nabla f_\sigma(y)^T.$$

Here

- f_σ is the convolution of the initial image f with a discretized Gaussian of zero mean and small standard deviation σ truncated between $[-3\sigma, 3\sigma]$ and mirrored at the boundary; this slight smoothing of the image before taking the discrete derivatives avoids too noisy gradients,
- ∇ denotes a discrete gradient operator, e.g., in this paper, forward differences in vertical and horizontal directions with mirror boundary condition, and
- the convolution of the rank-1 matrices $\nabla f_\sigma(y) \nabla f_\sigma(y)^T$ with the discrete Gaussian G_ρ of zero mean and standard deviation ρ truncated between $[-3\rho, 3\rho]$ is performed for every coefficient of the matrix; the choice of the parameter ρ usually results in a tradeoff between keeping too much noise in the image for small values and blurring the edges for large values.

Clearly, $\mathcal{J}_\rho(x)$ is a symmetric, positive semidefinite matrix. For natural or noisy images f it is in general positive definite, i.e. an element of $\mathcal{P}(2)$. Otherwise one could also exclude matrices with zero eigenvalues by choosing $\mathcal{V} \subset \mathcal{G}$ appropriately in our model (12).

Let $\lambda_1 \geq \lambda_2 > 0$ denote the eigenvalues of $\mathcal{J}_\rho(x)$, $x \in \mathcal{G}$, with corresponding normed eigenvectors v_1 and $v_2 = v_1^\perp$. Then $\lambda_1 \gg \lambda_2$, i.e., $\frac{\lambda_1}{\lambda_2} \gg 1$ indicates an edge at $x \in \mathcal{G}$. If the quotient is near 1 we have a homogeneous neighborhood of x or a vertex. In this paper, we are interested in edges. To this end, we follow an approach from [23] and consider

$$J_\rho(x) := \frac{1}{\sqrt{\det \mathcal{J}_\rho(x)}} \mathcal{J}_\rho(x).$$

Clearly, this matrix has the same eigenvectors as $\mathcal{J}_\rho(x)$ and the eigenvalues $\mu_1 = \sqrt{\frac{\lambda_1}{\lambda_2}}$ and $\mu_2 = 1/\mu_1$. Further, $\mu_1 \gg \mu_2$ indicates an edge. The matrix belongs to the subspace $\mathcal{P}_1(2) \subset \mathcal{P}(2)$ of symmetric positive definite matrices of determinant 1. The set $\mathcal{P}_1(2)$ together with the affine invariant metric of $\mathcal{P}(2)$ forms a Riemannian manifold which is isomorphic to the hyperbolic manifold of constant curvature $-\frac{1}{2}$, see Appendix C.

We consider images $(J_\rho(x))_{x \in \mathcal{G}}$ with values in $\mathcal{P}_1(2)$. The structure tensors are visualized as ellipses using the scaled eigenvectors $\mu_i v_i$, $i = 1, 2$ as major axes. The colorization is done with the anisotropy index from [58], i.e., tensors indicating edges are purple or blue, while red or green ellipses indicate constant regions.

We take the artificial image of size $N = M = 64$ with values in $[0, 1]$ shown in Fig. 6 (a) and add white Gaussian noise with standard deviation $\sigma_n = 0.2$ to the image. The noisy image is depicted in Fig. 6 (b). Fig. 6 (c) shows the structure tensor image J_{ρ_1} for $\sigma = 0.8$ and small $\rho_1 = \frac{35}{100}$. On the one hand, the structure tensors nicely show sharp edges, but on the other hand they are affected by a huge amount of noise.

Searching for a value of the parameter ρ_2 on a grid of size $\frac{1}{20}$ in order to find a structure tensors with thinnest edges and constant regions results in $\rho_2 = \frac{6}{5}$ and the structure tensors depicted in Fig. 6 (d). The larger parameter ρ increases the size of the neighborhood in the smoothing of the matrices $\nabla f_\sigma \nabla f_\sigma^T$. This yields smoother structure tensors, but the edges become broader (blurred). Some artifacts from the noise are still visible.

A remedy is shown in Fig. 6 (e), where we denoised the structure tensor image from Fig. 6 (c) using the PDRA applied to model (12) with the parameters $\alpha = 1.05$, $\eta = 0.4$,

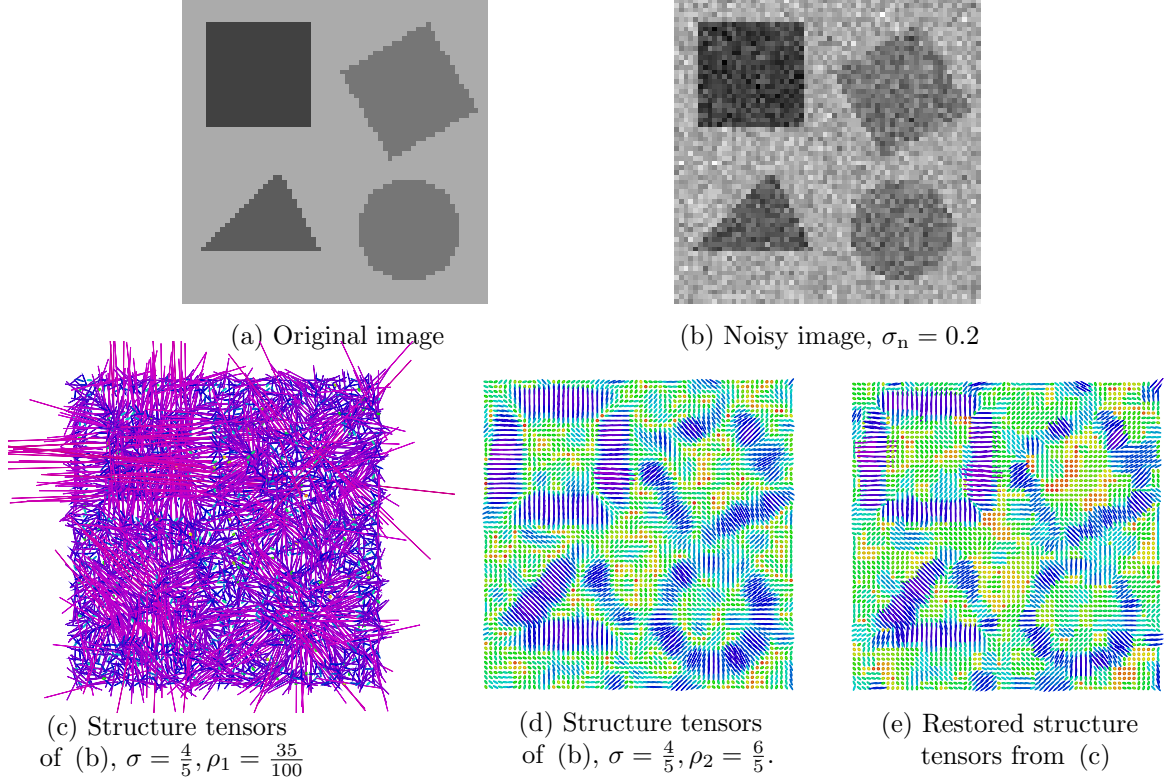


Figure 6. Application of model (12) and the PDRA to $\mathcal{P}_1(2)$ -valued images.

and $\lambda = 0.9$. The denoised structure tensor image shows thinner edges and less noise than the one in Fig. 6(d).

Note that the manifold-valued computations in the PDRA were done with respect to the hyperbolic manifold with pre-processing (post-processing) by the isomorphism (inverse isomorphism) from $\mathcal{P}_1(2)$ to the hyperbolic manifold described in Appendix C.

8.3. Denoising of Symmetric Positive Definite Matrices: DT-MRI

In magnetic resonance tomography, it is possible to capture the diffusivity of the measured material and obtain diffusion tensor images (DT-MRI), where each pixel is a 3×3 symmetric positive definite matrix. As already mentioned, the manifold $\mathcal{P}(3)$ is not of constant curvature, so that we cannot ensure the convergence of the PDRA in general. Nevertheless, we can investigate the convergence numerically. We compare the algorithm with the CPPA and HQMA. For the HQMA we chose two different parameters $\delta_1 := \frac{1}{10}$ and $\delta_2 := \frac{1}{100}$ in the smoothing function φ .

We process the Camino dataset²[26] which captures the diffusion inside a human head. To be precise, we take a subset from the complete dataset $f = (f_{i,j,k}) \in \mathcal{P}(3)^{112 \times 112 \times 50}$, namely from the traversal plane $k = 28$ the data points $(f_{i,j,28})$ with $(i, j) \in \{28, \dots, 87\} \times \{24, \dots, 73\}$. In the original Camino data some of the pixel are missing which we reset using a nearest neighbor approximation depicted in Fig. 7(b). The denoising result by applying

²see <http://cmic.cs.ucl.ac.uk/camino>

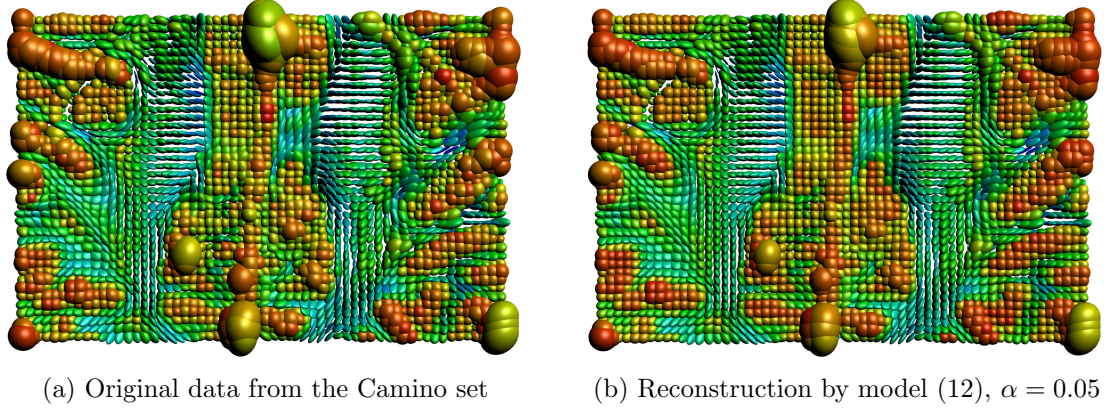


Figure 7. The Camino DT-MRI data of slice 28 and traversal plane subset $(i, j) \in \{28, \dots, 87\} \times \{24, \dots, 73\}$ (left) and its denoised version by model (12) with the PDRA.

	PDRA	CPPA	HQMA, δ_1	HQMA, δ_2
ℓ^2 -TV value	109.7844	109.7854	109.9617	109.7890
# iterations	49	1042	13	28
runtime (sec.)	139.2664	643.2270	113.0849	225.3906

Table 3. Comparison of all four algorithms to minimize the ℓ^2 -TV functional (12) with respect to the resulting minimal value, the number of iterations and the computational time.

the PDRA to model (12) with $\alpha = 0.05$ is shown in Fig. 7 (b). We have used the parameters $\eta = 0.58$, $\lambda = 0.93$, and $\epsilon = 10^{-6}$ within the algorithm. The other three algorithms yield similar results.

Again we compare the development of functional $\mathcal{E}^{(r)}$ and the error $\epsilon^{(r)}$, where we use $\epsilon = 10^{-6}$ as stopping criterion. The results are shown in Fig. 8. Both values decay fast in the first few iterations, so we plotted both with a logarithmic x axis. Comparing PDRA and CPPA, the results are similar to the experiments on the Retina data. Furthermore, the functional during the PDRA decreases below those of the HQMA, which is expected, because the HQMA is applied to the smoothed functional \mathcal{E}_δ . The error $\epsilon^{(r)}$ behaves quite similar in the PDRA and HQMA differing only by a factor.

The functional values, number of iterations and runtimes of the algorithms are presented in Table 3. While the HQMA with δ_1 is the fastest both in number of iterations and runtime, it is not able to reach the minimal functional value due to the smoothing. Even more, looking at a single iteration, the PDRA is faster than both HQMAs.

Though there is no mathematical proof of converging to a minimizer, the PDRA yields results, that still outperform the CPPA with respect to iterations and runtime. It is compatible with the HQMA, that minimizes a smoothed functional to gain performance.

8.4. Inpainting of Symmetric Positive Definite Matrices

In many situations, the measured data is not only affected by noise, but several data items are lost. In the artificial experiment shown in Fig. 9, the original image f on $\mathcal{G} = \{1, \dots, 16\}^2$

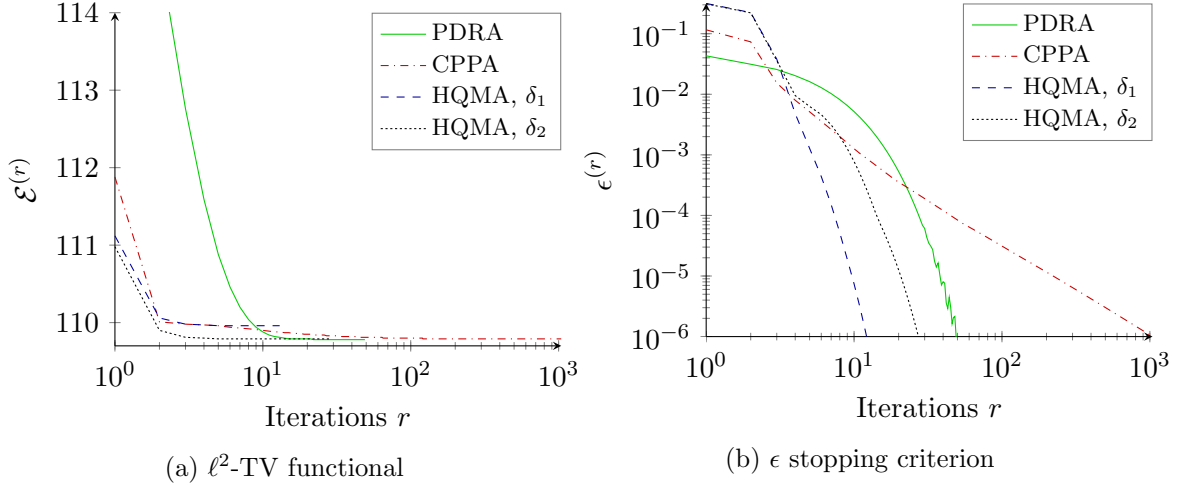


Figure 8. For the development of both the ℓ^2 -TV functional value in (a) and the stopping criterion $\epsilon^{(r)} = d(x^{(r)}, x^{(r-1)})$ in (b) we compare the CPPA, DR and HQ minimization, with two choices of δ . While the CPPA and the HQMA decrease the functional value a lot during the first iterations, the PDRA yields the minimal value first and the HQ minimization can not reach the minimum due to the relaxation. Regarding the $\epsilon^{(r)}$ the PDRA and HQMA behave similar, while the CPPA a slower decrease.

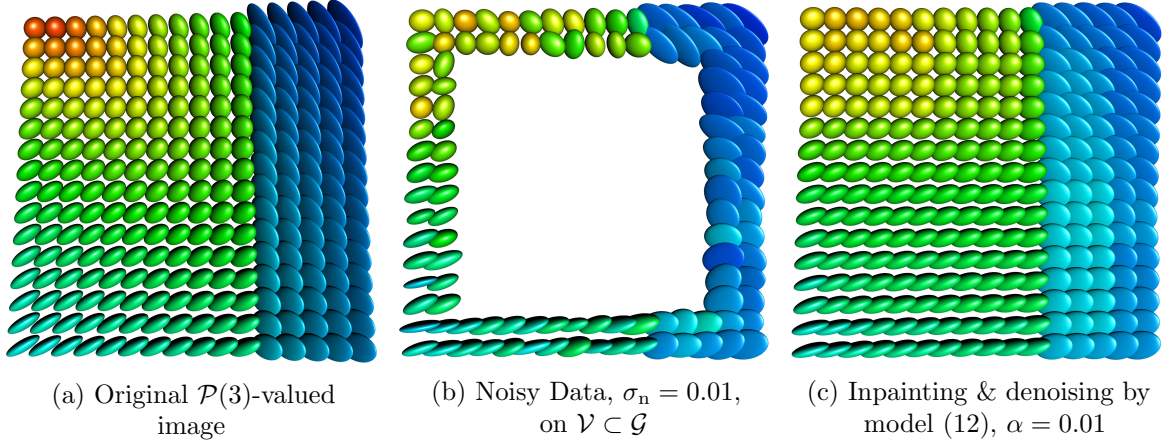


Figure 9. An artificial example of an $\mathcal{P}(3)$ -valued image, where (a) an original image is (b) obstructed by noise and loss of data. The reconstruction using the PDRA yields (c) a reconstruction that takes edges into account.

is destroyed in the inner part, i.e. only the indices $\mathcal{V} = \{(i, j) : \min\{i, 16 - i, j, 16 - j\} \leq 2\}$ are kept, and the remaining data is disturbed by Rician noise, $\sigma_n = 0.01$, cf. Fig. 9 (b). After initialization of the missing pixels by the nearest neighbor method, we employ the PDRA to minimize (12) with $\alpha = 0.1$, $\eta = 3$ and $\lambda = 0.95$. The algorithm stops after 117 iterations in 36.188 seconds with the stopping criterion $\epsilon = 10^{-5}$. The result is shown in Fig. 9 (c). We obtain $\mathcal{E}(u^{(117)}) = 5.7107$. The CPPA with $\eta = 12$ stops after 2210 iterations with the same stopping criterion as above in 161 seconds. The functional \mathcal{E} has a slightly higher value 5.7117.

In total, the PDRA requires far less iterations and a shorter computational time than the CPPA, even though one iteration of the PDRA takes four times as long as one iteration of the CPPA due to the computation of the Karcher mean. Nevertheless, the functional value of the PDRA also beats the CPPA when using the same stopping criterion. So even if we do not have a proof of convergence, the PDRA even performs better than the CPPA in case of noisy and lossy data.

9. Conclusions

We considered the restoration (denoising and inpainting) of images having values in symmetric Hadamard manifolds. We proposed a model with an L_2^2 data term, and anisotropic TV-like regularization term, and examined the performance of a parallel Douglas Rachford algorithm for minimizing the corresponding functional. Note that this carries over directly to an L_1 data term. Convergence can be proved for manifolds with constant non positive curvature. Univariate Gaussian probability distributions or symmetric positive definite matrices with determinant 1 are typical examples of such manifolds. Numerically, the algorithm works also well for other symmetric Hadamard manifolds as the symmetric positive definite matrices. However, having a look at the convergence proof, it is not necessary that the reflection operators are nonexpansive for arbitrary values on the manifold, but instead a fixed point is involved in the estimates. Can the convergence be proved under certain assumptions on the locality of the data? We are also interested in the question under which conditions the reflection operator a every proper convex lsc function on a Hadamard manifold with constant curvature is nonexpansive. We have seen that this is true for indicator functions of convex sets and various distance-like functions. For the latter, we have proved that the reflection is nonexpansive on general symmetric Hadamard manifolds. We are currently working on a software package for MATLAB which includes several algorithms to minimize ROF-like functionals involving first and second order differences applied to manifold-valued images. The code will be made public soon and the code is already available on request.

Acknowledgments We would like to thank M. Bačák for bringing reference [53] to our attention. Thanks to J. Angulo for providing the data of the Retina experiment. We gratefully acknowledge funding by the DFG within the project STE 571/13-1 & BE 5888/2-1.

A. Proof of Theorem 4.1

Proof. 1. Let $\hat{x} \in \mathcal{H}^n$ be a solution of (9). Then we obtain by Theorem 3.2 that

$$0 \in \partial(\varphi + \psi)(\hat{x}) = \partial\varphi(\hat{x}) + \partial\psi(\hat{x}).$$

This set inclusion can be split into two parts, namely there exists a point $y \in \mathcal{H}^n$ such that

$$\log_{\hat{x}} y \in \eta\partial\psi(\hat{x}) \quad \text{and} \quad -\log_{\hat{x}} y \in \eta\partial\varphi(\hat{x}),$$

i.e.,

$$0 \in \eta\partial\psi(\hat{x}) - \log_{\hat{x}} y \quad \text{and} \quad 0 \in \eta\partial\varphi(\hat{x}) + \log_{\hat{x}} y. \quad (32)$$

By Theorem 3.2 and since $\nabla d(y, \cdot)^2(\hat{x}) = -2\log_{\hat{x}} y$, we have

$$\tilde{x} = \text{prox}_{\eta\psi}(y) = \arg \min_{z \in \mathcal{H}^n} \left\{ \frac{1}{2}d(y, z)^2 + \eta\psi(z) \right\} \Leftrightarrow 0 \in -\log_{\tilde{x}} y + \eta\partial\psi(\tilde{x}). \quad (33)$$

Hence the first inclusion is equivalent to

$$\hat{x} = \text{prox}_{\eta\psi}(y). \quad (34)$$

This implies

$$\exp_{\hat{x}}(-\log_{\hat{x}} y) = R_{\hat{x}} y = \mathcal{R}_{\eta\psi}(y). \quad (35)$$

From the second inclusion in (32) we obtain

$$0 \in \eta\partial\varphi(\hat{x}) - \log_{\hat{x}}(\exp_{\hat{x}}(-\log_{\hat{x}} y)) = \eta\partial\varphi(\hat{x}) - \log_{\hat{x}}(\mathcal{R}_{\eta\psi}(y)). \quad (36)$$

Using again (33) this is equivalent to

$$\hat{x} = \text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y)). \quad (37)$$

Now (35) can be rewritten as

$$y = \exp_{\hat{x}}(-\log_{\hat{x}}(\mathcal{R}_{\eta\psi}(y)))$$

and plugging in (37) we get

$$\begin{aligned} y &= \exp_{\text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y))}(-\log_{\text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y))} \mathcal{R}_{\eta\psi}(y)) \\ &= \mathcal{R}_{\eta\varphi} \mathcal{R}_{\eta\psi}(y). \end{aligned}$$

Hence y is a fixed point of $\mathcal{R}_{\eta\varphi} \mathcal{R}_{\eta\psi}$ which is related to \hat{x} by (34).

2. Conversely, let y be a fixed point of $\mathcal{R}_{\eta\varphi} \mathcal{R}_{\eta\psi}$. Expanding the reflection on $\eta\varphi$, we obtain

$$y = \exp_{\text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y))}(-\log_{\text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y))} \mathcal{R}_{\eta\psi}(y)). \quad (38)$$

We set $\hat{x} := \text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y))$. Rewriting (38) yields

$$\exp_{\hat{x}}(-\log_{\hat{x}} y) = \exp_{\text{prox}_{\eta\psi}(y)}(-\log_{\text{prox}_{\eta\psi}(y)} y).$$

From the uniqueness of geodesics we get

$$\text{prox}_{\eta\varphi}(\mathcal{R}_{\eta\psi}(y)) = \hat{x} = \text{prox}_{\eta\psi}(y).$$

By (33) we conclude

$$0 \in \eta\partial\psi(\hat{x}) - \log_{\hat{x}} y,$$

and similar to (36) we have

$$0 \in \eta\partial\varphi(\hat{x}) + \log_{\hat{x}} y.$$

Adding these inclusions we obtain

$$0 \in \partial\varphi(\hat{x}) + \partial\psi(\hat{x}) \subseteq \partial(\varphi + \psi)(\hat{x}),$$

i.e., \hat{x} is a solution of (9). □

B. Symmetric Positive Definite Matrices $\mathcal{P}(n)$

The manifold $(\mathcal{P}(n), \langle \cdot, \cdot \rangle_{\mathcal{P}(n)})$ of symmetric positive definite $n \times n$ matrices $\mathcal{P}(n)$ is given by

$$\mathcal{P}(n) := \{x \in \mathbb{R}^{n \times n} : x = x^T \text{ and } \mathbf{a}^T x \mathbf{a} > 0 \text{ for all } \mathbf{a} \in \mathbb{R}^n\}$$

with the affine invariant metric

$$\langle u, v \rangle_{\mathcal{P}(n)} := \text{Trace}(x^{-1} u x^{-1} v), \quad u, v \in T_x \mathcal{P}(n).$$

We denote by Exp and Log the matrix exponential and logarithm defined by $\text{Exp } x := \sum_{k=0}^{\infty} \frac{1}{k!} x^k$ and $\text{Log } x := \sum_{k=1}^{\infty} \frac{1}{k} (I - x)^k$, $\rho(I - x) < 1$. Note that another metric for $\mathcal{P}(n)$, the so-called Log-Euclidean metric, was proposed in [5, 62] which is not considered in this paper.

Further, we use the following functions, see, e.g., [68]:

Geodesic Distance. The distance between two points $x, y \in \mathcal{P}(n)$ is defined as

$$d_{\mathcal{P}(n)}(x, y) := \|\text{Log}(x^{-\frac{1}{2}} y x^{-\frac{1}{2}})\|.$$

Exponential Map. The exponential map $\exp_x : T_x \mathcal{P}(n) \rightarrow \mathcal{P}(n)$ at a point $x \in \mathcal{P}(n)$ is defined as

$$\exp_x(v) := x^{\frac{1}{2}} \text{Exp}(x^{-\frac{1}{2}} v x^{-\frac{1}{2}}) x^{\frac{1}{2}}, \quad v \in T_x \mathcal{P}(n).$$

Logarithmic Map. The logarithmic map $\log_x : \mathcal{P}(n) \rightarrow T_x \mathcal{P}(n)$ at a point $x \in \mathcal{P}(n)$ is given by

$$\log_x(y) := x^{\frac{1}{2}} \text{Log}(x^{-\frac{1}{2}} y x^{-\frac{1}{2}}) x^{\frac{1}{2}}, \quad y \in \mathcal{P}(n).$$

Unit Speed Geodesic. The geodesic connecting x and y is given by

$$\gamma_{x,y}(t) := x^{\frac{1}{2}} \text{Exp}(t \text{Log}(x^{-\frac{1}{2}} y x^{-\frac{1}{2}})) x^{\frac{1}{2}}, \quad t \in [0, 1],$$

with $\gamma_{x,y}(0) = x$ and $\gamma_{x,y}(1) = y$.

C. Hyperbolic Space

In this section, we recall equivalent models of the hyperbolic space \mathbb{H}_M which were used in our implementations. We start with the hyperbolic manifold which is the basis of our computations and introduce the relevant manifold functions. Then we consider other equivalent models, namely the Poincaré ball, the Poincaré upper half-plane, the manifold of univariate Gaussian probability measures, and the space of symmetric positive definite matrices with determinant 1. In particular, we determine the isometries from these spaces to the hyperbolic manifold.

The *hyperbolic manifold* \mathbb{H}_M^d of dimension d can be embedded into the \mathbb{R}^{d+1} using the Minkowski inner product $\langle x, y \rangle_M := -x_{d+1}y_{d+1} + \sum_{i=1}^d x_i y_i$. Then

$$\mathbb{H}_M^d := \left\{ x \in \mathbb{R}^{d+1} : \langle x, x \rangle_M = -x_{d+1}^2 + \sum_{i=1}^d x_i^2 = -1, x_{d+1} > 0 \right\} \subseteq \mathbb{R}^{d+1},$$

together with the metric $g_M := \langle \cdot, \cdot \rangle_M$ is a Riemannian manifold. It has curvature -1 , see [50]. In this paper we are interested in $d = 2$. By \sinh and \cosh we denote the sine and cosine hyperbolicus and their inverses by arsinh and arcosh , respectively. The following functions were used in our computations:

Geodesic Distance. The distance between two points $x, y \in \mathbb{H}_M^d$ is defined as

$$d_{\mathbb{H}_M^d}(x, y) := \operatorname{arcosh}(-\langle x, y \rangle_M).$$

Exponential Map. The exponential map $\exp_x : T_x \mathbb{H}_M^d \rightarrow \mathbb{H}_M^d$ at a point $x \in \mathbb{H}_M^d$ is

$$\exp_x(v) := \cosh(\sqrt{\langle v, v \rangle_M})x + \sinh(\sqrt{\langle v, v \rangle_M}) \frac{v}{\sqrt{\langle v, v \rangle_M}}, \quad v \in T_x \mathbb{H}_M^d.$$

Logarithmic Map. The logarithmic map $\log_x : \mathbb{H}_M^d \rightarrow T_x \mathbb{H}_M^d$ at a point $x \in \mathbb{H}_M^d$ is given by

$$\log_x(y) := \frac{\operatorname{arcosh}(-\langle x, y \rangle_M)}{\sqrt{\langle x, y \rangle_M^2 - 1}} (y + \langle x, y \rangle_M x)$$

Geodesic. The geodesic connecting x and y reads

$$\begin{aligned} \gamma_{\widehat{x,y}}(t) &:= x \cosh(t \operatorname{arcosh}(-\langle x, y \rangle_M)) \\ &\quad + \frac{y + \langle x, y \rangle_M x}{\sqrt{\langle x, y \rangle_M^2 - 1}} \sinh(t \operatorname{arcosh}(-\langle x, y \rangle_M)), \quad t \in [0, 1], \end{aligned}$$

with $\gamma_{\widehat{x,y}}(0) = x$ and $\gamma_{\widehat{x,y}}(1) = y$.

Next we consider the other models together with the relevant bijections.

Poincaré ball (\mathbb{H}_B^d, g_B) . Let $B^d \subset \mathbb{R}^d$ be the unit ball with respect to the Euclidean distance. Together with the metric

$$g_B(u, v) := 4 \frac{\langle u, v \rangle}{(1 - \|x\|^2)^2}, \quad u, v \in T_x B^d,$$

it becomes a Riemannian manifold $\mathbb{H}_B^d \subset \mathbb{R}^d$ called the *Poincaré unit ball*. It is equivalent to the hyperbolic manifold \mathbb{H}_M^d with isometry $\pi_1: \mathbb{H}_M^d \rightarrow \mathbb{H}_B^d$ defined by

$$\pi_1(x) := \frac{1}{1 + x_{d+1}} \tilde{x}, \quad \pi_1^{-1}(y) = \frac{1}{1 - \|y\|^2} \begin{pmatrix} 2y \\ 1 + \|y\|^2 \end{pmatrix},$$

where $x = (x_1, \dots, x_d, x_{d+1})^T = (\tilde{x}^T, x_{d+1})^T$, see [50, Proposition 3.5].

Poincaré half-space (\mathbb{H}_P^d, g_P) . The upper half-space $\mathbb{H}_P^d := \{x \in \mathbb{R}^d : x_d > 0\}$. with the metric

$$g_P(u, v) := \frac{\langle u, v \rangle}{x_d^2}, \quad u, v \in T_x \mathbb{H}_P^d,$$

is a Riemannian manifold known as *Poincaré half-space*. It is equivalent to \mathbb{H}_B^d and thus to \mathbb{H}_M^d with isometry $\pi_2: \mathbb{H}_M^d \rightarrow \mathbb{H}_P^d$ given by

$$\pi_2(x) = \frac{1}{\|\tilde{x}\|^2 + (x_d - 1)^2} \begin{pmatrix} 2\tilde{x} \\ 1 - \|x\|^2 - x_d^2 \end{pmatrix}, \quad \pi_2^{-1}(y) = \frac{1}{\|\tilde{y}\|^2 + (y_d + 1)^2} \begin{pmatrix} 2\tilde{y} \\ \|\tilde{y}\|^2 + y_d^2 - 1 \end{pmatrix},$$

where $x = (x_1, \dots, x_{d-1}, x_d)^T = (\tilde{x}^T, x_d)^T$, see [50, Proposition 3.5].

Univariate Gaussian probability measures (\mathcal{N}, g_F) . A distance measure for probability distributions with density function $\varphi(x, \theta)$ depending on the parameters $\theta = (\theta_1, \dots, \theta_n)$ is given by the Fischer information matrix [72, Chapter 11]

$$F(\theta) = \left(\int_{-\infty}^{\infty} \varphi(x, \theta) \frac{\partial \ln \varphi(x, \theta)}{\partial \theta_i} \frac{\partial \ln \varphi(x, \theta)}{\partial \theta_j} dx \right)_{i,j=1}^n.$$

In the case of univariate Gaussian densities it reduces to

$$F(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

Then the Fischer metric is given by

$$g_F(u, v) := \frac{u_1 v_1 + 2u_2 v_2}{\sigma^2}, \quad u, v \in T_{(\mu, \sigma)} \mathcal{N},$$

i.e. $ds^2 = u^T F u = \frac{1}{\sigma^2} (u_1^2 + 2u_2^2)$. Then, the isomorphism $\pi_3: \mathcal{N} \rightarrow \mathbb{H}_P^2$ with

$$\pi_3(\mu, \sigma) = \left(\frac{\mu}{\sqrt{2}}, \sigma \right), \quad \pi_3^{-1}(x_1, x_2) = (\sqrt{2}x_1, x_2)$$

is an isometry between (\mathcal{N}, g_F) and $(\mathbb{H}_P^2, 2g_P)$, see, e.g., [27]. Thus, (\mathcal{N}, g_F) has curvature $-\frac{1}{2}$. We mention that $\pi_1^{-1} \circ \pi_2^{-1} \circ \pi_3$ is an isomorphism from \mathcal{N} to our model hyperbolic manifold \mathbb{H}_M^2 .

Symmetric positive definite 2×2 matrices with determinant 1 $(\mathcal{P}_1(2), \langle \cdot, \cdot \rangle_{\mathcal{P}_1(2)})$. Following [23], an isomorphism $\pi_3: \mathcal{P}_1(2) \rightarrow \mathbb{H}_M^2$ is given by

$$\pi_4(a) := \begin{pmatrix} \frac{a_{11}-a_{22}}{2} \\ a_{12} \\ \frac{a_{11}+a_{22}}{2} \end{pmatrix}, \quad \pi_4^{-1}(x) = \begin{pmatrix} x_1 + x_3 & x_2 \\ x_2 & x_3 - x_1 \end{pmatrix},$$

where

$$a := \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \quad \text{with} \quad \det a = a_{11}a_{22} - a_{12}^2 = 1,$$

and $x := (x_1, x_2, x_3)^T$ with $x_3 > 0$ and $x_1^2 + x_2^2 - x_3^2 = -1$.

The operator π_3 is an isometry between $(\mathcal{P}_1(2), \langle \cdot, \cdot \rangle_{\mathcal{P}_1(2)})$ and $(\mathbb{H}_M^2, 2\langle \cdot, \cdot \rangle)$. To verify this relation, we define the push-forward operator $\phi_*: T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$ of a smooth mapping $\phi: \mathcal{M} \rightarrow \mathcal{N}$ between two manifolds \mathcal{M}, \mathcal{N} as

$$\phi_*u(f) := u(f \circ \phi),$$

where $f \in \mathbb{C}^\infty(\mathcal{N}, \mathbb{R})$, $u \in T_x\mathcal{M}$. The push-forward ϕ_* of ϕ is also known as the differential of ϕ , for details see [51].

Lemma C.1. *Let $(\mathbb{H}_M^2, 2\langle \cdot, \cdot \rangle_M)$, the model space $\mathcal{M}_{-\frac{1}{2}}^2$ embedded into the \mathbb{R}^3 , and $(\mathcal{P}_1(2), \langle \cdot, \cdot \rangle_{\mathcal{P}})$, the symmetric positive definite 2×2 matrices having determinant 1 with the affine invariant metric, be given. Then*

$$\psi: \mathbb{H}_M^2 \rightarrow \mathcal{P}_1(2), \psi(x) = \pi_4^{-1}(x) = \begin{pmatrix} x_1 + x_3 & x_2 \\ x_2 & x_3 - x_1 \end{pmatrix},$$

is an isometry.

Proof. We have to show that

$$2\langle u, u \rangle_M = \langle \psi_*(u), \psi_*(u) \rangle_{\psi(x)},$$

for all $u \in T_x\mathbb{H}_M^2$, see [50]. The push-forward ψ_* is given by

$$\psi_*(u) = \begin{pmatrix} u_1 + u_3 & u_2 \\ u_2 & u_3 - u_1 \end{pmatrix}.$$

Given $x \in \mathbb{H}_M^2$ and $u \in T_x\mathbb{H}_M^2$ we know

$$\langle x, x \rangle_M = x_1^2 + x_2^2 - x_3^2 = -1, \tag{39}$$

$$\langle x, u \rangle_M = x_1u_1 + x_2u_2 - x_3u_3 = 0. \tag{40}$$

Further we have

$$(\psi(x))^{-1} = \begin{pmatrix} x_3 - x_1 & -x_2 \\ -x_2 & x_1 + x_3 \end{pmatrix}.$$

We start with

$$\begin{aligned}
\langle \psi_*(u), \psi_*(u) \rangle_{\psi(x)} &= \text{trace} \left((\psi(x))^{-1} \psi_*(u) (\psi(x))^{-1} \psi_*(u) \right) \\
&= \text{trace} \left(\left(\begin{pmatrix} x_3 - x_1 & -x_2 \\ -x_2 & x_1 + x_3 \end{pmatrix} \begin{pmatrix} u_1 + u_3 & u_2 \\ u_2 & u_3 - u_1 \end{pmatrix} \right)^2 \right) \\
&= \text{trace} \left(\begin{pmatrix} x_3 u_1 + x_3 u_3 - x_1 u_1 - x_1 u_3 - x_2 u_2 & x_3 u_2 - x_1 u_2 - x_2 u_3 + x_2 u_1 \\ x_3 u_2 + x_1 u_2 - x_2 u_3 - x_2 u_1 & -x_2 u_2 + x_1 u_3 - x_1 u_1 + x_3 u_3 - x_3 u_1 \end{pmatrix} \right)^2 \\
&= 2(x_1^2 u_1^2 - x_1^2 u_2^2 + x_1^2 u_3^2 - x_2^2 u_1^2 + x_2^2 u_2^2 + x_2^2 u_3^2 + x_3^2 u_1^2 + x_3^2 u_2^2 + x_3^2 u_3^2 \\
&\quad + 4x_1 x_2 u_1 u_2 - 4x_1 x_3 u_1 u_3 - 4x_2 x_3 u_2 u_3) \\
&= 2(u_1^2(x_1^2 - x_2^2 + x_3^2) + u_2^2(-x_1^2 + x_2^2 + x_3^2) + u_3^2(x_1^2 + x_2^2 + x_3^2) \\
&\quad + 4x_1 x_2 u_1 u_2 - 4x_1 x_3 u_1 u_3 - 4x_2 x_3 u_2 u_3).
\end{aligned}$$

Using (39) we obtain

$$\begin{aligned}
\langle \psi_*(u), \psi_*(u) \rangle_{\psi(x)} &= 2(u_1^2(2x_1^2 + 1) + u_2^2(2x_2^2 + 1) + u_3^2(2x_3^2 - 1) \\
&\quad + 4x_1 x_2 u_1 u_2 - 4x_1 x_3 u_1 u_3 - 4x_2 x_3 u_2 u_3) \\
&= 2(u_1^2 + u_2^2 - u_3^2) \\
&\quad + 2(u_1^2 x_1^2 + u_2^2 x_2^2 + u_3^2 x_3^2 + 4x_1 x_2 u_1 u_2 - 4x_1 x_3 u_1 u_3 - 4x_2 x_3 u_2 u_3) \\
&= 2\langle u, u \rangle_M + 2(u_1 x_1 + u_2 x_2 - u_3 x_3)^2.
\end{aligned}$$

With (40) we obtain the assertion. \square

References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton and Oxford, Princeton University Press, 2008.
- [2] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.
- [3] A. D. Aleksandrov. A theorem on triangles in a metric space and some of its applications. *Trudy Matematicheskogo Instituta V.A. Steklova*, 38:5–23, 1951.
- [4] J. Angulo and S. Velasco-Forero. Morphological processing of univariate Gaussian distribution-valued images based on Poincaré upper-half plane representation. In *Geometric Theory of Information*, pages 331–366. Springer, 2014.
- [5] V. Arsigny, X. Pennec, P. Fillard, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.

- [6] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24(3):1542–1566, 2014.
- [7] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. De Gruyter, Berlin, 2014.
- [8] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. A second order non-smooth variational model for restoring manifold-valued images. *SIAM Journal on Scientific Computing*, 38(1):A567–A597, 2016.
- [9] P. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
- [10] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.
- [11] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *Journal of the Optical Society of America A*, 19:1334–1345, 2002.
- [12] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Hybrid projection-reflection method for phase retrieval. *Journal of the Optical Society of America A*, 20(6):1025–1034, 2003.
- [13] H. H. Bauschke and W. M. Moursi. On the order of the operators in the Douglas-Rachford algorithm. *Preprint arXiv: 1505.02796*, 2015.
- [14] R. Bergmann, R. H. Chan, R. Hielscher, J. Persch, and G. Steidl. Restoration of manifold-valued images by half-quadratic minimization. *Inverse Problems in Imaging*, accepted and *Preprint arXiv: 1505.07029*, 2015.
- [15] R. Bergmann, F. Laus, G. Steidl, and A. Weinmann. Second order differences of cyclic data and applications in variational denoising. *SIAM Journal on Imaging Sciences*, 7(4):2916–2953, 2014.
- [16] R. Bergmann and A. Weinmann. Inpainting of cyclic data using first and second order differences. In *EMCVPR2015*, Lecture Notes in Computer Science, pages 155–168, Berlin, 2015. Springer.
- [17] J. Borwein and B. Sims. The Douglas–Rachford algorithm in the absence of convexity. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 93–109. Springer, 2011.
- [18] J. M. Borwein and M. K. Tam. A cyclic Douglas–Rachford iteration scheme. *Journal of Optimization Theory and Applications*, 160(1):1–29, 2014.
- [19] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-positive Curvature*, volume 319. Springer Science & Business Media, 1999.
- [20] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A generic approach to the filtering of matrix fields with singular PDEs. In *Scale Space and Variational Methods in Computer Vision*, pages 556–567. Springer, 2007.

- [21] B. Burgeth, M. Welk, C. Feddern, and J. Weickert. Morphological operations on matrix-valued images. In *Computer Vision - ECCV 2004*, Lecture Notes in Computer Science, 3024, pages 155–167, Berlin, 2004. Springer.
- [22] C. Ched'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In *Computer Vision—ECCV 2002*, pages 251–265. Springer, 2002.
- [23] P. Chossat and O. Faugeras. Hyperbolic planforms in relation to visual edges and textures perception. *PLoS Computational Biology*, 5(12):e1000625, 2009.
- [24] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), 2008.
- [25] P. L. Combettes and J.-C. Pesquet. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. *Preprint arXiv: 1404.7536*, 2014.
- [26] P. A. Cook, Y. Bai, S. Nedjati-Gilani, K. K. Seunarine, M. G. Hall, G. J. Parker, and D. C. Alexander. Camino: Open-source diffusion-MRI reconstruction and processing. In *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 2759, Seattle, WA, USA, 2006.
- [27] S. I. R. Costa, S. A. Santos, and J. E. Strapasson. Fisher information distance: a geometrical reading. *Discrete Applied Mathematics*, 197:59–69, 2015.
- [28] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956.
- [29] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293 – 318, 1992.
- [30] V. Elser, I. Rankenburg, and P. Thibault. Searching with iterated maps. *Proceedings of the National Academy of Sciences*, 104(2):418–423, 2007.
- [31] J. H. Eschenburg. Lecture notes on symmetric spaces. *Preprint*, 1997.
- [32] A. Fernández-León and A. Nicolae. Averaged alternating reflections in geodesic spaces. *Journal of Mathematical Analysis and Applications*, 402(2):558 – 566, 2013.
- [33] O. P. Ferreira and P. R. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51(2):257–270, 2002.
- [34] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, 1987.
- [35] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, chapter IX, pages 299–340. North-Holland, Amsterdam, 1983.

- [36] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Computer and Mathematics with Applications*, 2:17–40, 1976.
- [37] M. Giaquinta and D. Mucci. The BV-energy of maps into a manifold: relaxation and density results. *Annali della Scuola Normale Superiore di Pisa – Classe di Scienze*, 5(4):483–548, 2006.
- [38] M. Giaquinta and D. Mucci. Maps of bounded variation with values into a manifold: total variation and relaxed energy. *Pure Applied Mathematics Quarterly*, 3(2):513–538, 2007.
- [39] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer, New York, 1984.
- [40] R. Glowinski. On alternating direction methods of multipliers: a historical perspective. In *Modeling, Simulation and Optimization for Science and Technology*, pages 59–82. Springer, 2014.
- [41] S. Gravel and V. Elser. Divide and concur: A general approach to constraint satisfaction. *Physical Review E*, 78(3):036706, 2008.
- [42] P. Grohs and J. Wallner. Interpolatory wavelets for manifold-valued data. *Applied and Computational Harmonic Analysis*, 27(3):325–333, 2009.
- [43] P. Grohs and J. Wallner. Definability and stability of multiscale decompositions for manifold-valued data. *Journal of the Franklin Institute*, 349(5):1648–1664, 2012.
- [44] S. Helgason. *Geometric Analysis on Symmetric Spaces*. AMS, 1994.
- [45] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM Journal on Optimization*, 23(4):2397–2419, 2013.
- [46] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *Preprint arXiv: 1410.1390*, 2014.
- [47] J. Jost. *Nonpositive Curvature: Geometric and Analytic Aspects*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1997.
- [48] B. Kakavandi. Weak topologies in complete CAT(0) metric spaces. *Proceedings of the American Mathematical Society*, 141(3):1029–1039, 2013.
- [49] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.
- [50] J. M. Lee. *Riemannian Manifolds. An Introduction to Curvature*. Springer-Verlag, New York-Berlin, New York-Berlin-Heidelberg, 1997.
- [51] J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer Science & Business Media, New York, NY, 2012.

- [52] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In *IEEE ICCV 2013*, pages 2944–2951, 2013.
- [53] C. Li, B. S. Mordukhovich, J. Wang, and J.-C. Yao. Weak sharp minima on Riemannian manifolds. *SIAM Journal on Optimization*, 21(4):1523–1560, 2011.
- [54] G. Li and T. K. Pong. Global convergence of splitting methods for nonconvex composite optimization. *Preprint arXiv: 1407.0753*, 753, 2014.
- [55] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two linear operators. *SIAM Journal on Numerical Analysis*, 16:964–976, 1979.
- [56] S. Magnússon, P. C. Weeraddana, M. G. Rabbat, and C. Fischione. On the convergence of alternating direction lagrangian methods for nonconvex structured optimization problems. *Preprint arXiv: 1409.8033*, 2014.
- [57] U. Mayer. Gradient flows on nonpositively curved metric spaces and harmonic maps. *Communications in Analysis and Geometry*, 6(2):199–253, 1998.
- [58] M. Moakher and P. G. Batchelor. Symmetric positive-definite matrices: From geometry to applications and visualization. In *Visualization and Processing of Tensor Fields*, pages 285–298. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [59] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace Hilbertien. *Comptes Rendus de l’Académie des Sciences de Paris, Série a*, 255:2897–2899, 1962.
- [60] M. Nikolova and M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, 27(3):937–966, 2005.
- [61] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 72:383–390, 1979.
- [62] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006.
- [63] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schröder. Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation*, 4(4):1201–1232, 2005.
- [64] J. G. Rešetnjak. Non-expansive maps in a space of curvature no greater than K . *Sibirskii Matematičeskii Žurnal*, 9:918–927, 1968.
- [65] G. Rosman, X.-C. Tai, R. Kimmel, and A. M. Bruckstein. Augmented-Lagrangian regularization of manifold-valued maps. *Methods and Applications of Analysis*, 21(1):105–122, 2014.
- [66] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1):259–268, 1992.
- [67] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 92(3):265–280, 2011.

- [68] S. Sra and R. Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *Preprint arxiv:1320.1039v3*, 2014.
- [69] G. Steidl, S. Setzer, B. Popilka, and B. Burgeth. Restoration of matrix fields by second order cone programming. *Computing*, 81:161–178, 2007.
- [70] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDs. *SIAM Journal on Numerical Analysis*, 42(2):686–713, 2004.
- [71] E. Strekalovskiy and D. Cremers. Total variation for cyclic structures: Convex relaxation and efficient minimization. In *IEEE CVPR 2011*, pages 1905–1911. IEEE, 2011.
- [72] J. A. Thomas and T. M. Cover. *Elements of Information Theory*, volume 2. Wiley New York, 2006.
- [73] C. Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.
- [74] J. Wallner, E. N. Yazdani, and P. Grohs. Smoothness properties of Lie group subdivision schemes. *Multiscale Modeling & Simulation*, 6(2):493–505, 2007.
- [75] F. Wang, W. Cao, and Z. Xu. Convergence of multi-block bregman ADMM for non-convex composite problems. *Preprint arXiv: 1505.03063*, 2015.
- [76] Y. Wang, W. Yin, and J. Zeng. Self equivalence of the alternating direction method of multipliers. *Preprint arXiv: 1511.06324*, 2015.
- [77] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.
- [78] J. Weickert, C. Feddern, M. Welk, B. Burgeth, and T. Brox. PDEs for tensor image processing. In *Visualization and Processing of Tensor Fields*, pages 399–414, Berlin, 2006. Springer.
- [79] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, 2014.
- [80] M. Welk, C. Feddern, B. Burgeth, and J. Weickert. Median filtering of tensor-valued images. In *Pattern Recognition*, Lecture Notes in Computer Science, 2781, pages 17–24, Berlin, 2003. Springer.
- [81] Y. Xu, W. Yin, Z. Wen, and Y. Zhang. An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers of Mathematics in China*, 7(2):365–384, 2012.
- [82] M. Yan and W. Yin. Self equivalence of the alternating direction method of multipliers. *Preprint arXiv: 1407.7400*, 2014.
- [83] M. Yan and W. Yin. Monotone vector fields and the proximal point algorithm on Hadamard manifolds. *Preprint arXiv: 1407.7400v3*, 2015.